AD_____

Award Number: DAMD17-02-1-0578

TITLE: Characterization of the Novel DNA-Binding Activity of p270, a hSWI/SNF Protein Frequently Downregulated in Breast Cancer

PRINCIPAL INVESTIGATOR: Antonia Patsialou
Elizabeth Moran, Ph.D.

CONTRACTING ORGANIZATION: Temple University
Philadelphia, PA 19140

REPORT DATE: July 2005

TYPE OF REPORT: Annual Summary

20060223 008

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01-07-2005	Annual Summary	1 Jul 2004 – 30 Jun 2005

4. TITLE AND SUBTITLE

Characterization of the Novel DNA-Binding Activity of p270, a hSWI/SNF Protein Frequently Downregulated in Breast Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
DAMD17-02-1-0578

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Antonia Patsialou
Elizabeth Moran, Ph.D.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Temple University
Philadelphia, PA 19140

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Human SWI/SNF complexes are ATP-dependent chromatin remodelers that play fundamental roles in the regulation of cell growth, development and tumor suppression. p270 is an integral member of these complexes and is absent in approximately 10% of breast tumors. The ARID region is the most prominent motif of the p270 protein and it is important for the function of the protein *in vitro*. This suggests that this region has an important role in the tumor suppressor function of p270. The ARID is a DNA-binding motif that is diagnostic of a family of proteins that are important in development, tissue specific gene expression and tumorigenesis. My studies therefore concentrated on the ARID-dependent DNA-binding properties of p270. Through a combination of structural, biochemical and mutational approaches, valuable data about the structural integrity of the domain and its interaction with DNA have emerged. This biochemical information can be important for drug design or the development of diagnostic/prognostic tools in breast cancer. Furthermore, this biochemical analysis will be a very useful tool in the literature for future studies on the physiological role of p270 and its tumor suppressor functions in the human SWI/SNF complexes.

15. SUBJECT TERMS
DNA-binding domains, site-specific mutagenesis

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Antonia Patsialou
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	UU	82	19b. TELEPHONE NUMBER (include area code) 215-707-7312

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

Introduction

Human SWI/SNF complexes are ATP-dependent chromatin remodelers that play fundamental roles in the regulation of cell growth, development and tumor suppression (Vignàli *et al.*, 2001, Martens *et al.*, 2003). p270/ARID1A is an integral member of hSWI/SNF complexes (Dallas *et al.*, 1998). Decreased expression or complete absence has been reported in various tumor cell lines for almost all of the subunits of the complexes, including p270 (e.g. Wong *et al.*, 2000, Roberts *et al.*, 2000, DeCristofaro *et al.*, 2001, Muchardt and Yaniv, 2001). Our lab has obtained tissue array data suggesting that p270/ARID1A is deficient in approximately 10% of breast tumors and 30% of kidney tumors (Wang *et al.*, 2004), as well as recent results that indicate p270 is required for normal cell cycle arrest in differentiating cells (Nagl *et al.*, in revision). The most prominent motif of the p270 protein is its DNA-binding motif, the ARID, suggesting that this region contributes to the tumor suppressor function of p270-containing SWI/SNF complexes. The ARID is a highly structured helix-turn-helix motif-based domain, which is conserved in all studied eukaryotes, and is diagnostic of a family that includes at least 15 distinct human proteins. ARID family proteins play important roles in development, tissue specific gene expression, and tumor suppression (reviewed in Wilsker *et al.*, 2002 and 2005). The first ARID proteins identified bind preferentially to AT-rich sequences (Gregory *et al.*, 1996, Herrscher *et al.*, 1995). However, the ARID-dependent DNA-binding activity of p270 is sequence-independent (Dallas *et al.*, 2000, Wilsker *et al.*, 2004). We now know, through studies supported partly by this fellowship, that most ARID proteins bind DNA in a sequence-independent manner like p270 (Patsialou *et al.*, 2005).

When I applied for this fellowship, few biochemical studies had addressed the DNA-binding properties attributed to the ARID region, and no detailed mutational characterization of any ARID domain had been reported. The ARID DNA-binding domain is required for the physiological function of at least one of the sequence specific members of the ARID family of proteins (Shandala *et al.*, 2002). It is not yet known whether the ARID is required for the physiological function of p270, but it is important in an *in vitro* transactivation assay (Nie *et al.*, 2000). My application focused on a detailed characterization of the ARID region of p270 through a combination of structural, biochemical, and mutational approaches. In this final report, I show a summary of the major findings and outcomes of the work supported by this fellowship.

Body

My application had three objectives:

1. To determine whether p270 can convert native DNA into a form more able to enhance the ATPase activity of BRG1.
2. To characterize p270 DNA-binding activity in human breast cell lines.
3. To extend my structural studies of the ARID region of p270.

TASK 3: I have completed Task 3. Several results of experiments performed as part of this task were shown in the previous two progress reports. Additional results, as well as the methods used to generate mutants tested in DNA-binding assays, are described in detail in two manuscripts published by our lab (Wilsker *et al.*, 2004, Patsialou *et al.*, 2005). Copies of these manuscripts are attached in the appendix of this final report. Throughout my studies, our lab communicated with Dr. Yuan Chen and her group at the Beckman Research Institute of the City of Hope in California, who has worked on the NMR structure of the p270 ARID region. Structural studies for the ARID region have been reported in the literature for three more ARID family proteins. Based on the knowledge generated by these structural studies of the domain and sequence homologies among the ARID domains, I generated substitution mutants targeting both highly conserved aromatic/hydrophobic residues throughout the domain and basic/polar residues in the predicted areas of DNA interaction. Overall, I have made 3 single substitution mutants in the Dri ARID (one of the first AT-rich specific ARID proteins identified and one of four for which the structure has been solved) and 12 single substitution mutants and 17 multiple substitution mutants in p270 (from which 8 were made at the time of my application). This substitution analysis in the p270 ARID region gave us important insight on the general properties of the ARID DNA-binding. First, an aromatic scaffold is important for the structural integrity of the ARID. Specifically, an invariant tryptophan is indispensable for the right folding of the domain in both p270 and Dri. The two proteins, though, seem to also have a difference in their tolerance of substitutions in their aromatic scaffolds. When an invariant tyrosine is substituted with an alanine, the p270 ARID DNA-binding is comparable to the wild-type, while the one in Dri is defective. I therefore hypothesized that distinct ARID proteins differ in the inherent flexibility of their domains, a

property that is probably reflected by their broad range of specificity properties (some of the domains are AT-rich specific and some are sequence-independent). This is discussed further in Patsialou *et al.*, 2005. Second, I have identified residues in the p270 ARID that are important for contact with DNA. These contacts are basic or polar residues in areas that were predicted to be important from previous NMR studies of two other ARID proteins, Dri and MRF2 (Iwahara *et al.*, 2002, Zhu *et al.*, 2001). The biochemical data agree with the structural study of the p270 ARID in complex with DNA (Kim *et al.*, 2004), but additionally I identified the specific residues involved in contacting the DNA in these areas. Finally, combinations of substitutions in more than one area of the domain showed that these residues act synergistically for the correct orientation and the interaction of the domain with the DNA. These studies, in combination with the structural data in the literature, provide detailed information about the properties of individual amino acids in the ARID, and the general properties of the DNA-binding activity of the domain. Overall, my study provides useful information that may ultimately be important in drug design and diagnostic/prognostic screens, as it could identify mutants that are functionally impaired in breast cells.

TASK 1:   In this task I had proposed to determine whether the ARID has a physiological role in the p270 function in the SWI/SNF complex. In the original application, I proposed to address this by exploring the manner in which DNA-binding might contribute to the activity of the complex. In previous reports I have shown that p270 binds preferentially to single-stranded DNA, and I hypothesized that p270 can convert native DNA into a form potentially more able to enhance the ATPase activity of BRG1. BRG1 is the DNA-dependent ATPase and thus the motor of the SWI/SNF complexes. In my first year, I cloned and expressed a bacterial peptide that contained the ATPase domain of BRG1. In that year's report I included preliminary results showing that this peptide had indeed ATPase activity. In my second year of work, I overcame many difficulties in purifying the enzyme, however I still could not establish storage conditions that maintained stability of the enzyme. Stability was essential in this task because I needed to use the enzyme in multiple biochemical assays in order to test the contribution of p270. In my initial application I had also proposed to use immuno-precipitated complexes from p270 deficient lines for the same purposes. I explored this approach in my second year as well, but due to inability to completely separate cellular DNA from the purified complex, the *in vivo* assay was not sufficiently DNA-dependent to reveal preference in the form of DNA provided.

Therefore, in my second year report, I developed a new plan for testing the requirement of the ARID region for the p270 function. As described in more detail in that report, a normal pre-osteoblast cell line was developed in our lab with small interfering RNA (siRNA) technology that has decreased expression of p270 (for a review in siRNAs see Brummelkamp et al., 2002). Reduction in the expression levels of p270 results in deficient differentiation-associated cell cycle arrest (Nagl *et al.*, in revision). The p270 knock-down lines failed to express the differentiation marker Alkaline Phosphatase (ALP) when induced to differentiate into osteoblasts. They also had a different expression pattern of cell cycle regulating proteins when compared to the parental line (examples of these results are shown in the second year report). This gave us a very useful assay for testing the function of p270 in cells, and the role of the ARID in p270 function. I had therefore proposed to re-introduce p270 in these cell-lines and assay for restoration of cell cycle controls. Then, I would introduce a p270 construct with the ARID region deleted (p270ΔARID) and test if the ARID deletion abrogates the restoration. The constructs that were used in these experiments had silent mutations designed specifically to allow the peptides to escape silencing by the siRNA in the knock-down lines. I had proposed to re-introduce p270 either by retrovirus transduction or by generation of stable lines (again this hypothesis and the procedures were described in more detail in my second year report). In this third year of my work, I used both these approaches to re-introduce the p270 and p270ΔARID constructs in the knock-down lines. For the retroviral transduction, I used the phoenix system (see http://www.stanford.edu/group/nolan/retroviral_systems/phx.html). For the stables lines, a plasmid vector with a puromycin resistance gene was co-transfected with the p270 constructs to allow for clone selection. After screening about 80 independent lines of each construct for expression of the exogenous peptides by western blot, I found three independent lines expressing the p270 peptide and two independent lines expressing the p270ΔARID peptide. Surprisingly, the expression levels for both constructs were extremely low (Figure 1). The levels of expression were indeed as low as the levels of the downregulated endogenous p270 in these knock-down lines. The same results were seen with the retrovirus approach, as well.

As described in my second year report, our lab had preliminary data that these constructs can express the p270 peptides in relatively high levels when co-tranfected with the specific small interfering RNAs used for the knock-down lines (seen in lanes 1 and 4 of Figure 1). The difference in these preliminary data and the stable lines could result from the fact that the first was a transient tranfection. Presumambly the mutations give enough immunity to escape silencing

temporarily in a transient transfection, but not in a stable genomic integration system such as a stable line. Because of the low levels of expression, we were unable to test the contribution of the ARID in the p270 function at this time. Our lab is currently exploring other systems by which p270 can be re-introduced in the knock-down lines (such as an adenovirus system) and test the contribution of the ARID in the tumor suppression function of the complexes.
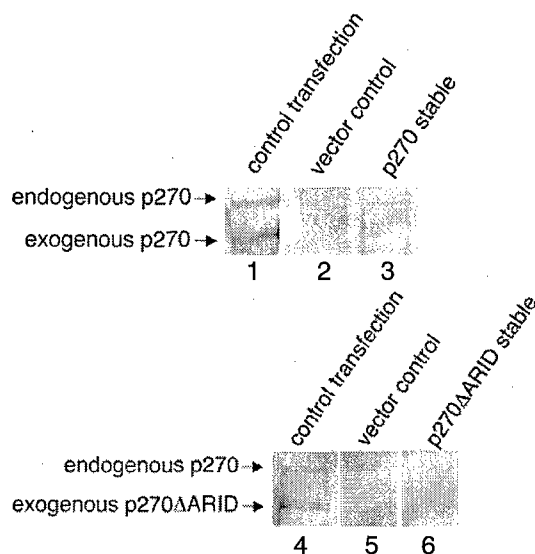

Figure 1. Expression of the p270 and p270ΔARID peptides in the knock-down lines.
One of the independent stable cell lines produced for each construct and a control line produced with empty vector are shown in this figure (lanes 3, 6 and 2, 5 respectively). Total cell lysate from the cells was run in a 7% SDS-polyacrylamide gel and then probed for p270 expression by western blot using the PSG3 antibody previously described in our lab (Dallas *et al.*, 1998). The exogenous p270 peptides are missing 377 amino acids in the N-terminus (out of the total 2,285 amino acids of the full-length protein), and thus run smaller in the gel than the endogenous protein. Transient expression of the same constructs in 293 cells is shown as a control (lanes 1 and 4). 293 cells are of kidney epithelial origin and have normal levels of endogenous p270.

TASK 2:   In this task, I had proposed to use the knowledge produced by the results from the above mentioned aims to characterize the DNA-binding properties of p270 specifically in human breast cancer cell lines. I had proposed to use the *in vivo* ATPase assay of the immuno-precipitated complexes and the characterization of the requirement of the ARID in the cell cycle regulating functions of p270 in order to test the properties of the p270 protein in breast cancer lines. Unfortunately, as described above, we were unable to test these assays as of now due to technical difficulties. Other physiological experiments concerning the role of p270 in cell cycle regulation are underway in our lab (e.g. Nagl *et al.*, in revision). These experiments will give us very important information on the role and function of p270 as a tumor suppressor.

Key Research Accomplishments
- Completed a detailed mutagenesis and biochemical analysis of the p270 ARID region
- Established that the ARID region is stabilized by an aromatic scaffold
- Established that there are differences in p270 and Dri in how they tolerate substitutions in their aromatic scaffold
- Identified residues in p270 that are important for DNA contact, potentially useful as diagnostic tools in breast cancer
- Showed that different areas of the domain can act synergistically for the correct orientation and the interaction of the domain with DNA
- Participated in writing and publishing a review article about the ARID family of proteins
- Participated in writing and publishing two research articles about the DNA-binding properties of the ARID
- Currently preparing one review article on the role of p270/ARID1A and the SWI/SNF complexes on tumorigenesis
- Currently participating on the writing and publishing of a research manuscript on the role of p270 in cell cycle arrest
- Presented a poster with the results of this fellowship at the Department of Defense Era of Hope Breast Cancer Research Program Conference in Philadelphia, June 2005
- Results from this work have been used in applications for funding in our lab by my advisor Dr. E. Moran

Reportable outcomes
Degrees:
With the support of this fellowship, I have completed the requirements of my degree. I am graduating this summer with a Ph.D. in Molecular Biology and Genetics from Temple University School of Medicine and the Fels Institute for Cancer Research and Molecular Biology.
Reviews:
- D. Wilsker, A. Patsialou, P.B. Dallas and E. Moran. 2002. *ARID proteins: A diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development.* Cell Growth & Differentiation (Review) 13: 95-106.
- A. Patsialou, N.G. Nagl, Jr., D. Wilsker, X. Wang and E. Moran. *Human SWI/SNF complexes and control of cell growth and differentiation.* International Review of Cytology: A Survey of Cell Biology. Academic Press, Inc., San Diego. Invited review. In preparation. IRC reviews are published as hard-bound volumes covering 35-50 printed pages. This volume is scheduled for publication in the Summer of 2005.

Publications:
- Wilsker D, Patsialou A, Zumbrun SD, Kim S, Chen Y, Dallas PB, Moran E. 2004. *The DNA-binding properties of the ARID-containing subunits of yeast and mammalian SWI/SNF complexes.* Nucleic Acids Res. 24;32(4):1345-53.
- Patsialou A, Wilsker D, Moran E. *Sequence-specificity properties of the ARID family.* Nucleic Acids Res. 33: 66-80.
- N.G. Nagl Jr, A. Patsialou, D.S. Haines, P.B. Dallas, G.R. Beck Jr. and E. Moran E. *The p270 (ARIDA/SMARCF1) subunit of mammalian SWI/SNF-related complexes is essential for normal cell cycle arrest.* In revision.

Abstracts:
- A. Patsialou, D. Wilsker and E. Moran. *Characterization of the DNA-binding activity of p270, a human SWI/SNF complex subunit frequently downregulated in breast cancer.* Presented at the Department of Defense Era of Hope Breast Cancer Research Program Conference, Philadelphia PA, June 2005
- NG. Nagl. Jr., A Patsialou, S Flowers, GR. Beck, Jr., PB. Dallas, and E Moran. *Functional complementation between adenovirus E1A targets and the p270 subunit of SWI/SNF-related complexes.* Selected for oral presentation at the at the Molecular Biology of DNA Tumor Viruses Conference in Madison, WI July, 2004.
- D Wilsker, A Patsialou, X Wang, PB. Dallas and E Moran. *The DNA-binding properties of p270 (SMARCF1/ARID1A), an integral member of the human SWI/SNF complexes.* Poster presentation at the 94th AACR meeting, July 2003, Washington DC.
- A Patsialou, S Kim, M van Scoy, PB. Dallas, Y Chen and E Moran. *Mutagenesis analysis of the ARID region of p270, a human SWI/SNF complex protein.* Selected for oral presentation at the Fels Institute Annual Research Day, May 2003.
- A Patsialou, S Kim, M van Scoy, PB. Dallas, Y Chen and E Moran. *p270 DNA binding in relation to ARID protein family structure.* Poster presentation at the Dawn Marks Annual Research Day at the Temple University School of Medicine, December 2002.
- D Wilsker, A Patsialou, PB. Dallas and E Moran. *DNA-binding properties of p270, a member of the ARID family of DNA-binding proteins implicated in nuclear hormone receptor signaling.* Era of Hope Department of Defense Breast Cancer Research Program Meeting, September 2002, Orlando, FL.

Conclusions

This fellowship focused on the study of the DNA-binding properties of the p270 ARID region. p270 is a member of the human SWI/SNF tumor suppressor complexes and is frequently downregulated in breast cancer. The ARID region is the most prominent motif of the protein and it is important for the function of the protein *in vitro,* suggesting an important role for the DNA-binding activity in the tumor suppressor function of p270. My studies resulted in valuable data about the structural integrity of the domain as well as its interaction with DNA. As discussed above, the information from the point mutations in the domain can be important for drug design or the development of diagnostic/prognostic tools. Furthermore, the biochemical analysis can be a very useful tool for future studies concerning the physiological role of p270 in the SWI/SNF tumor suppressor functions.

References

- Brummelkamp TR, Bernards R, Agami R. 2002. *A system for stable expression of short interfering RNAs in mammalian cells.* Science. 296(5567): 550-3

- Dallas, P. B., I. W. Cheney, D. Liao, V. Bowrin, W. Byam, S. Pacchione, R. Kobayashi, P. Yaciuk, and E. Moran. 1998. *p300/CREB binding protein-related protein p270 is a component of mammalian SWI/SNF complexes.* Mol. Cell. Biol. 18: 3596-3603

- Dallas, P. B., S. Pacchione, D. Wilsker, V. Bowrin, R. Kobayashi, and E. Moran. 2000. *The human SWI/SNF complex protein, p270, is an ARID family member with nonsequence -specific DNA binding activity.* Mol. Cell Biol. 20: 3137-3146

- Decristofaro MF, Betz BL, Rorie CJ, Reisman DN, Wang W, Weissman BE. 2001. *Characterization of SWI/SNF protein expression* in *human breast cancer cell lines and other malignancies.* J Cell Physiol. 186(1): 136-4

- Gregory, S. L., R. D. Kortschak, B. Kalionis, and R. Saint. 1996. *Characterization of the dead ringer gene identifies a novel, highly conserved family of sequence-specific DNA binding proteins.* Mol. Cell. Biol. 16: 792-799

- Herrscher, R. F., M. H. Kaplan, D. L. Lelsz, C. Das, R. Scheuermann, and P. W. Tucker. 1995. *The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family.* Genes Dev. 9: 3067-3082

- Iwahara, J., M. Iwahara , G. W. Daughdrill, J. Ford, and R. T. Clubb. 2002. *The structure of the Dead ringer-- DNA complex reveals how AT-rich interaction domains (ARIDs) recognize DNA.* EMBO J. 21(5): 1197-1209

- Kim S, Zhang Z, Upchurch S, Isern N, Chen Y. 2004. *Structure and DNA-binding sites of the SWI1 AT-rich interaction domain (ARID) suggest determinants for sequence-specific DNA recognition.* J Biol Chem. 279(16): 16670-6

- Martens JA, Winston F. 2003. *Recent advances in understanding chromatin remodeling by Swi/Snf complexes.* Curr Opin Genet Dev. 13(2): 136-42

- Roberts CWM, Galusha SA, McMenamin ME, Fletcher CDM, Orkin SH. 2000. *Haploinsuffiency of Snf5 (integrase interactor 1) predisposes to malignant rhabdoid tumors in mice.* PNAS. 97(25): 13796-13800

- Shandala T, Kortschak RD, Saint R. 2002. *The Drosophila retained/dead ringer gene and ARID gene family function durig development.* Int J Dev Biol. 46(4): 423-430

- Vignali M, Hassan AH, Neely KE, Workman JL. 2000. *ATP-dependent chromatin-remodeling complexes.* Mol Cell Biol. 20(6): 1899-910

- Wang X, Nagl NG Jr, Flowers S, Zweitzig D, Dallas PB, Moran E. 2004. *Expression of p270 (ARID1A), a component of human SWI/SNF complexes in human tumors.* Int J Cancer. 112(4): 636-642

- Wilsker D., Patsialou A, Dallas PB, Moran E. 2002. *ARID proteins: A diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development.* Cell Growth & Differentiation (Review). 13: 95-106

- Wilsker D, Patsialou A, Zumbrun SD, Kim S, Chen Y, Dallas PB, Moran E. 2004. *The DNA-binding properties of the ARID-containing* subunits *of yeast and mammalian SWI/SNF complexes.* Nucleic Acids Res. 32(4): 1345-53

- Wilsker D, Probst L, Wain HM, Maltais L, Tucker PW, Moran E. 2005. *Nomenclature of the ARID family of DNA-binding proteins.* Genomics. 86(2): 242-51

- Wong AKC, Shanahan F, Chen Y, Lian L, Ha P, Hendricks K, Ghaffari S, Iliev D, Penn B, Woodland A, Smith R, Salada G, Carillo A, Laity K, Gupte J, Swedlund B, Tavtigian SV, Teng DHF, Lees E. 2000. *BRG1, a component of the SWI/SNF complex, is mutated in multiple human tumor cell lines.* Cancer Res. 60(21): 6171-6177

- Zhu, L., J. Hu, D. Lin, R. Whitson, K. Itakura, and Y. Chen. 2001. *Dynamics of the Mrf-2 DNA binding domain free and in complex with DNA.* Biochemistry. 40: 9142-50.

DNA-binding properties of ARID family proteins

Antonia Patsialou, Deborah Wilsker and Elizabeth Moran*

Fels Institute for Cancer Research and Molecular Biology, Temple University School of Medicine, Philadelphia PA 19140, USA

ABSTRACT

The ARID (*A–T Rich Interaction Domain*) is a helix–turn–helix motif-based DNA-binding domain, conserved in all eukaryotes and diagnostic of a family that includes 15 distinct human proteins with important roles in development, tissue-specific gene expression and proliferation control. The 15 human ARID family proteins can be divided into seven subfamilies based on the degree of sequence identity between individual members. Most ARID family members have not been characterized with respect to their DNA-binding behavior, but it is already apparent that not all ARIDs conform to the pattern of binding AT-rich sequences. To understand better the divergent characteristics of the ARID proteins, we undertook a survey of DNA-binding properties across the entire ARID family. The results indicate that the majority of ARID subfamilies (i.e. five out of seven) bind DNA without obvious sequence preference. DNA-binding affinity also varies somewhat between subfamilies. Site-specific mutagenesis does not support suggestions made from structure analysis that specific amino acids in Loop 2 or Helix 5 are the main determinants of sequence specificity. Most probably, this is determined by multiple interacting differences across the entire ARID structure.

INTRODUCTION

The ARID (*A–T Rich Interaction Domain*) is a helix–turn–helix motif-based DNA-binding domain, conserved in all eukaryotes and diagnostic of a family that comprises 15 distinct human proteins. ARID proteins, although diverse in function, all appear to play important roles in development, tissue-specific gene expression and cell growth regulation [reviewed in (1,2)]. The ARID consensus sequence, which spans about 100 residues, was first identified as a DNA-binding domain in the mouse B cell-specific transcription factor, Bright (3), and in the Dead ringer protein (Dri) of *Drosophila melanogaster* (4). Dri and Bright were each isolated in searches designed to detect proteins binding selectively to AT-rich sequences. Recognition of the Bright/Dri consensus defined the parameters of a new DNA-binding domain, and the properties of Bright and Dri inspired its name. MRF-1 and MRF2, which bind the CMV enhancer and repress its activity, are also ARID-containing proteins that bind selectively to AT-rich sites (5,6).

Although the first studied ARID-containing proteins bind preferentially to AT-rich sites, this behavior does not appear to be an intrinsic feature of the domain. Most ARID family members have not been characterized with respect to their DNA-binding behavior, but it has become apparent that not all ARIDs conform to the pattern of binding AT-rich sequences. p270 is a human ARID-containing protein that is an integral member of SWI/SNF-related chromatin remodeling complexes (7–10). p270 contains a complete ARID consensus and binds DNA with an affinity similar to Dri, but is unable to select oligonucleotides of any preferred sequence from a random pool (9,11). Lack of sequence specificity has been shown independently for the ARID family member, Osa, the closest *Drosophila* counterpart of p270 (12).

The ARID structures for MRF2, Dri, p270 and its yeast counterpart SWI1 have been studied by NMR (13–16). Despite the high degree of conservation in the domain, at least three distinct structural patterns are recognized: MRF2 and SWI1 both have six helices and two loops. Dri has one more helix on each end formed by sequences outside the consensus and a β-sheet instead of a flexible loop between Helix 1 and Helix 2. p270 has an additional short N-terminal helix, but no C-terminal helix or any β-sheets. The structures of the MRF2, Dri and p270 ARIDs have also been solved in complex with DNA (15,17,18). All studies agree that the ARID binds DNA via both the major and the minor grooves, and that major groove contacts are made through residues in Loop 2 and/or Helix 5.

The human ARID family can be divided into seven subfamilies based on the degree of sequence identity between individual members (Figures 1 and 2). The diverse

---

*To whom correspondence should be addressed. Tel: +215 707 7313; Fax: +215 707 6989; Email: betty@temple.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

```
            H0              H1         Loop1 or β-sheet    H2      H3      H4     Loop2    H5      H6              H7
ARID1A  SSSSTTTNEKITKLYELGGEPERKHWVDRYLAFTEEKAMGNTNLPAVGRKPLDLYRLYVSVKEIGGLTQVNKNKKWRELATNLNVG-TSSSAASSLKKQYIQCLYAFECKIER
ARID1B  SSSSTTTGEKITKVYELGNEPERKLWVDRYLTFMEERGSPVSSLPAVGKKPLDLFRLYVCVKEIGGLAQVNKNKKWRELATNLNVG-TSSSAASSLKKQYIQYLFAFECKIER

ARID2         MANSTGKAPPDERRKGLAFLDELRQFHHSRGSPFKKIPAVGGKELDLHGLYTRVTTLGGFAKVSEKNQWGEIVEEFNFPRSCSNAAFALKQYYLRYLEKYEKVHHF

ARID3A  DWTYEEQFK---QLYELDGDPKRKEFLDDLFSFMQKRGTPVNRIPIMAKQVLDLFMLYVLVTEKGGLVEVINKKLWREITKGLNLPTSITSAAFTLRTQYMKYLYPYECEKRGLSNPNELQAAIDSNRREGRRQ
ARID3B  GGRGREISRDFAKLYELDGDPKRKEFLDDLFVFMQKRGTPINRIPIMAKQILDLYMLYKLVTEKGGLVEIINKKIWREITKGLNLPTSITSAAFTLRTQYMKYLYAYECEKKALSSPAELQAAIDGNRREGRRP
ARID3C    MTPILPQLYELDADPKRKEFLDDLFSFMQKRGTPVNRVPIMAKQVLDLYALFRLVTAKGGLVEVINRKVWREVTRGLSLPTTITSAAFTLRTQYMKYLYPYECETRALSSPGELQAAIDSNRREGRRQ

ARID4A  EKEKEAKKTEEEVPEEELDPEERDNFLQQLYKFMEDRGTPINKPVLGYKDLNLFKLFRLVYHQGGCDNIDSGAVWKQIYMDLGIPILNSAASYNLKTAYRKYLYGFEEYCRS
ARID4B  DEKEKEDNSSEEEBEIEPFPEBRENFLQQLYKFMBDRGTPINKRPVLGYRNLNLFKLFRLVHKLGGFDNIESGAVWKQVYQDLGIPVLNSAAGYNVKCAYKKYLYGFEEYCRS

ARID5A  PISLEDSPEAGGEREEEQEREEEQAFLVSLYKFMKERHTPIERVPHLGFKQINLWKIYKAVEKLGAYELVTGRRLWKNVYDELGGSPGSTSAATCTRRHYERLVLPYVRHLKG
ARID5B  KVSNEEKPKVAIG---BECRADEQAFLVALYKYMKERKTPIERIFYLGFKQINLWTMFQAAQKLGYETTTARRQWKHIYDELGGNPGSTSAATCTRRHYERLILPYERFIK

JARID1A  VKSFRFTPRVQRLNELEAMTRVRLDFLDQLAKFWELQGS-TLKIPVVERKILDLYALSKIVASKGGFEHVTKEKKWSKVGSRLGYLPGK-GTGSLLKSHYERILYPYELFQSG
JARID1B  VDKLHFTPRIQRLNELEAQTRVKLNFLDQIAKYWELQGS-TLKIPHVERKILDLFQLNKLVAEEGGPAVVCKDRKWTKIATKMGFAPGK-AVGSHIRGHYERILNPYNLFLSG
JARID1C  VDNFRFTPRIQRLNELEAQTRVKLNYLDQIAKFWEIQGS-SLKIPNVERRILDLSLKIVVEEGGYEAICKDRRWARVAQRLNYPPGK-NIGSLLRSHYERIVYPYEMYQSG
JARID1D  VDNFRFTPRVQRLNELEAQTRVKLNYLDQIAKFWEIQGS-SLKIPNVERKILDLYSLSKIVIEEGGYEAICKDRRWARVAQRLHYPPGK-NIGSLLRSHYERIIYPYEMFQSG

JARID2  NDEMRFVTQIQHIHKLGRRWGPNVQRLACIKKHLKSQGITHDELPLIGGCELDLACFFRLINEHGGMQQVTELKKWNKLSDMLRIPKTAQERLAKLQEAYCQYILSYDSLSPE

Dri     GWSFEEQFKQVRQLYEINDDPKRKEFLDDLFSFMQKRGTPINRLPIMAKSVLDYELYNLVIARGGLVDVINKKLWQETIKGLHLPSSITSAAFTLRTQYMKYLYPYECEKKNLSTPAELQAAIDGNRREGRRS
SWI1          NNKQYELFMKSLIENCKKRNMPLQSIPEIGNRKINYFLYHLVQKFGGADQVTRTQQWSMVAQRLQI-----SDYQQLESIFRILLPYERHMIS
```
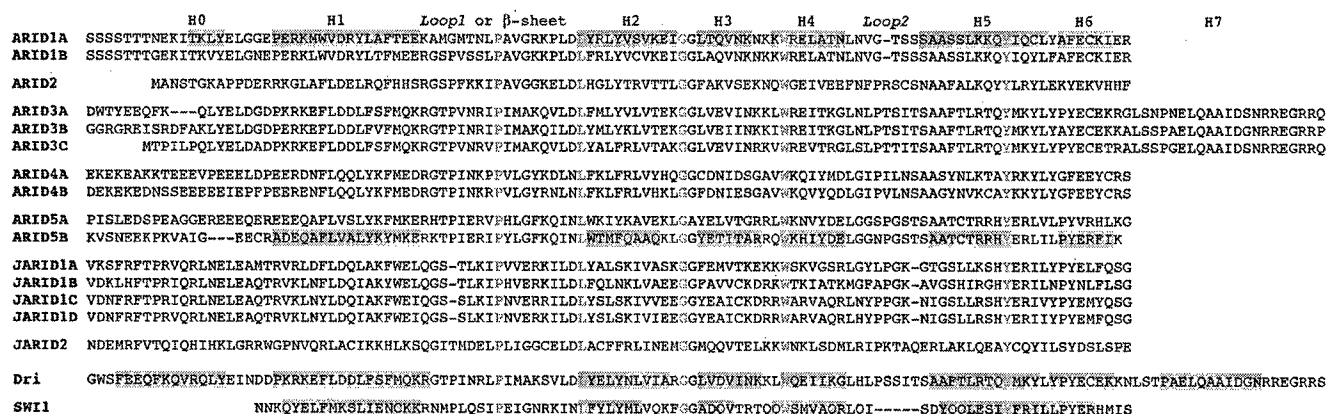
Figure 1. Alignment of human ARID sequences. The amino acid sequences of the ARID region of the 15 human ARID family members are shown. The shading indicates the boundaries of the α-helices where the structural data are known (13,15,17). The sequences of the ARIDs of *D.melanogaster* Dri, which has an ARID3-class sequence and *S.cerevisieae* SWI1, are shown as well because structural data is also available for them (14,16,18). Helices are labeled at the top from H0 to H7. The location of Loops 1 and 2 and the β-sheet (which so far is found only in the ARID3-class sequence) are also shown. The five invariant residues of the ARID region are shown in red. Part of the 'extended ARID' that characterizes the ARID3 subfamily is shown to indicate the degree of homology in this region. The ARID2 and ARID3C sequences begin at the initial methionine of the protein. The sequences were aligned using the Clustal W 1.8 multiple sequence alignment program (50). The computer-generated alignment was modified slightly to reflect higher level structural data.
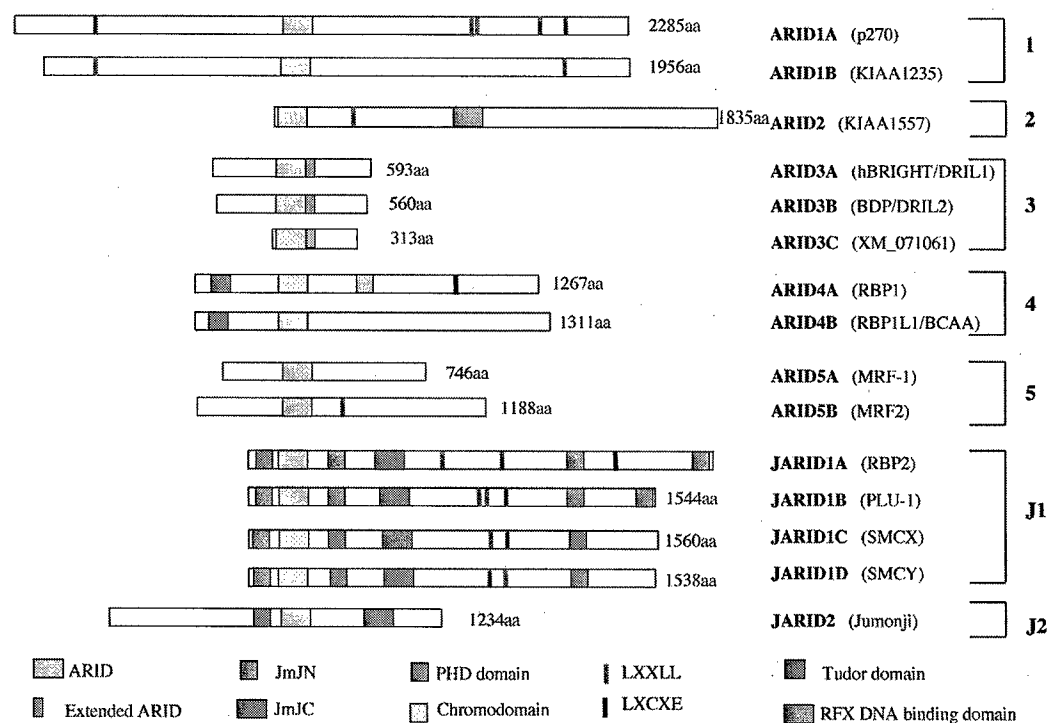
```
┌──────────────────────────────────────┐ 2285aa   ARID1A  (p270)        ┐
└──────────────────────────────────────┘                                ├ 1
┌──────────────────────────────────────┐ 1956aa   ARID1B  (KIAA1235)    ┘

      ┌─────────────────────────────────┐ 1835aa  ARID2  (KIAA1557)     ─ 2

        ┌───────────────┐ 593aa     ARID3A  (hBRIGHT/DRIL1)  ┐
        ┌───────────────┐ 560aa     ARID3B  (BDP/DRIL2)      ├ 3
        ┌───────┐ 313aa             ARID3C  (XM_071061)      ┘

      ┌─────────────────────────┐ 1267aa   ARID4A  (RBP1)          ┐
      ┌─────────────────────────┐ 1311aa   ARID4B  (RBP1L1/BCAA)   ┘ 4

      ┌──────────────────┐ 746aa    ARID5A  (MRF-1)   ┐
      ┌────────────────────────┐ 1188aa  ARID5B  (MRF2)    ┘ 5

    ┌──────────────────────────────┐          JARID1A  (RBP2)     ┐
    ┌──────────────────────────────┐ 1544aa   JARID1B  (PLU-1)    │ J1
    ┌──────────────────────────────┐ 1560aa   JARID1C  (SMCX)     │
    ┌──────────────────────────────┐ 1538aa   JARID1D  (SMCY)     ┘

   ┌───────────────────┐ 1234aa   JARID2  (Jumonji)   ─ J2
```

ARID   JmJN   PHD domain   LXXLL   Tudor domain
Extended ARID   JmJC   Chromodomain   LXCXE   RFX DNA binding domain

Figure 2. The human ARID family of proteins. Genome sequencing reveals 15 ARID-containing proteins in humans. The ARID family proteins can be grouped into subfamilies based on their similarity to each other within the ARID domain. The nomenclature described here reflects this subclassification of the family and clarifies their relationships to each other. A subset of ARID-containing proteins also contains JmJN and JmJC domains, and the proposed nomenclature reflects these relationships as well. Within the proposed subgroups of the ARID family, members typically have 70–85% identity within their ARID sequences, while across subgroups identity within the ARID sequence drops to ~25–30%. The 15 human ARID family proteins are represented by open bars and are aligned according to the position of the ARID sequence (indicated in yellow). The relative positions of other well-characterized domains and motifs are represented by differently colored bars or boxes in the appropriate protein structures and identified at the bottom of the figure. The amino acid (aa) length of each protein is shown to the right of the bar. The presence of additional motifs was identified through the Pfam database (51).

characteristics of the ARID proteins studied so far prompted a survey of DNA-binding properties across the entire ARID family. The results indicate that the majority of ARID subfamilies (i.e. five out of seven) bind DNA without obvious sequence preference. DNA-binding affinity also varies somewhat between subfamilies. Site-specific mutagenesis does not support suggestions made from structure analysis that specific amino acids in Loop 2 or Helix 5 are the main

determinants of sequence specificity. Most probably, this is determined by multiple interacting differences across the entire ARID structure.

MATERIALS AND METHODS

Plasmids

GST fusion constructs. The p270 fusion protein experiments were originally performed with the product of plasmid pNDX (9). For the mutagenesis studies, a shorter expression construct designated pNDB8 was generated to be more comparable in size with the Dri fusion peptide. The NDB8 plasmid expresses amino acids 958–1188 of p270 (numbering is according to the sequence at accession number NP_006006). The Dri fusion protein is the product of p410 (4), which was kindly provided by R. Saint (University of Adelaide, Australia) and expresses Dri residues 258–410 (according to accession number AAB05771).

A series of plasmids was assembled expressing GST-fusion proteins containing the ARID regions of representative members of ARID subfamilies. The MRF2 fusion protein is the product of pMRF2-GST, which was constructed by ligating a BamHI/SalI restriction fragment from the insert of plasmid MRF2pQE30 [(13); kindly provided by Yuan Chen at the Beckman Institute, City of Hope, Duarte, CA] into the pGEX4T vector (Pharmacia Biotech). A construct containing the ARID domain of human RBP1, called GST–ARID (19), was provided Dr Philip Branton (McGill University, Montreal, Canada). The ARID2 sequence was generated by RT–PCR from HepG2 cells using oligonucleotides ARID2-F (5′-ATAATGGCAAACTCGACGGGGAAG) and ARID2-R (5′-CACCCCGGCATTAGCAAGTAGTAA) to yield a 630 bp fragment that encodes amino acids 1–209 according to accession number XP_350876. The fragment was cloned into pCR2.1-TOPO vector (Invitrogen) to make ARID2-TOPO. The EcoRI fragment of ARID2-TOPO was sub-cloned into the EcoRI site of pGEX-4T1 (Pharmacia Biotech) to make pARID2-pGEX. The PLU-1 sequence was generated by RT–PCR from MCF-7 cells using oligonucleotides PLU-1 For (5′-TTCGCGGACCCCTTCGCTTTCA) and PLU-1 Rev short (5′-AATATTCATGGCCTCTGCTCTC). The reaction generated a 597 bp fragment extending from nucleotide 213 to 810 (according to accession number AJ132440.1), which was cloned into the pCR2.1-TOPO vector to create pPLU-1-TOPO. A PLU-1 containing BstXI restriction fragment was released from the vector, blunted with T4 DNA polymerase, and ligated with SmaI digested pGEX-4T1 to generate pPLU-1-GST. pPLU-1-GST generates a GST-fusion protein containing amino acids 42–241 of PLU-1 according to accession number CAB43532. An RBP2 sequence-containing PCR fragment was generated with primers RBP2-F-Xho (5′-AGACTCGA-GTTCACAGATCCGCTCAGCTTTATC) and RBP2-R-Xho (5′-AGACTCGAGTTTAGGACACCTCCAGTCTCCTTT) from the plasmid template pCMV-HA-RBP2 (provided by Philip Branton), and cloned into the pCR2.1-TOPO vector to create pRBP2-TOPO. An XhoI restriction fragment from the RBP2-TOPO insert was blunted with Klenow polymerase and ligated with SmaI-digested pGEX-4T3 to create the plasmid pRBP2-pGEX. This construct produces a GST-fusion protein containing RBP2 amino acids 29–339 (accession

number NP_005047). The jumonji fragment was amplified by PCR from a murine brain cDNA library in a vector backbone of pACT-2 (Clontech) that was kindly provided by Dr Premkumar Reddy (Fels Institute, Temple University School of Medicine, Philadelphia, PA). jumonji-TOPO was generated by PCR using the primers jumonji For (5′-AGAGAATTCTGTGAAAATCGTTCTACCTCGCAA) and jumonji Rev (5′-AGACTCGAGATGACAGTCCTTCTCTT-CCACTAA) to generate a 1030 bp fragment extending from nucleotide 1750 to 2780 according to accession number D31967. The PCR fragment was cloned into the pCR2.1-TOPO vector to create pjumonji-TOPO, excised from the vector with EcoRI and XhoI, and cloned into EcoRI/XhoI-digested pGEX-4T1 to create pjumonji-pGEX-4T1. This construct creates a GST-fusion protein containing amino acids 519–858 of jumonji (accession number NP_068678). This is the only case where a murine sequence was used in the subfamily constructs, but the mouse and human proteins are 92% identical across the coding span of the insert.

In vitro translation constructs. The p270 pNE9-B2 *in vitro* expression plasmid and the Dri *in vitro* expression plasmid pDriT2 were described previously (11).

Generation of amino acid substitution mutations

All mutations were generated using the QuikChange (Stratagene) system according to the manufacturer's instructions. The forward primers used to generate amino acid substitutions in pNE9-B2 or pDriT2 are as follows (the substituted bases are underlined):

p270.P1042A: GCATGACAAATCTGGCTGCTGTGGG-TAGGAAACC
p270.W1073A: GGTCAACAAGAACAAAAAAGCGCG-GGAACTTGCAACC
p270.Y1096A: CCTTGAAAAAGCAGGCTATCCAGTG-CTCTATGC
Dri.P306A: CCGATCAATCGGCTGGCGATAATGGCC-AAATCGG
Dri.W337A: CAACAAGAAGCTGGCGCAGGAGATCA-TCAAGGGGC
Dri.Y361A: CCCTGCGCACCCAAGCCATGAAGTATC-TGTACCCG

The remaining mutations were generated in p410 or pNDB8. The forward primers used to generate amino acids substitutions are as follows (the substituted bases are underlined):

Dri.SSS: GCCCTCCAGCATCTCCAGTGCCGCCTCCT-CCCTGCGCACCC
p270.TFT: GTGGGCACATCAACCAGTGCTGCCTTCA-CCTTGAAAAAGCAG

Deletion mutants were generated by a loop-out technique using a primer designed to form a junction between residues at the borders of the deletion. The sequence of the forward primers used to generate the deletions is as follows (the nucleotides that mark the boundaries of the loop-out are underlined):

Dri.ΔC: GAATCTGAGCACGCAGATGCCGATGACG
p270.ΔN: GGGACACCCAAGACAGAAATCACCAAGT-TGTATGAGCTG

Chimera mutants were made by first looping-out the sequence to be replaced, and then looping-in the desired sequence. The sequence of the forward primers used to generate the deletions is as follows (the nucleotides that mark the boundaries of the loop-out are underlined):

L2.H5.out: CGGGAACTTGCAACCAACCTCTTGAAA-AAGCAGTATATCCAG
H4.out: GGATTGACTCAGGTCAACAAGAACAAACT-CCACCTGCCCTCCAGC

The sequence of the forward primers used to generate the insertions is as follows (the nucleotides that form the loop-in are underlined):

L2.H5.in: GCAACCAACCTCCACCTGCCCTCCAGCA-TCACCAGTGCCGCCTTCACCTTGAAAAAGCAG
H4.in: GTCAACAAGAACAAACTGTGGCAGGAGAT-CATCAAGGGGCTCCACCTGCC

The sequence changes and the integrity of the surrounding sequences for all mutants were verified by DNA sequencing.

Sequence-specific selection of DNA

GST-fusion proteins were used in pull-down assays with a pool of Lambda DNA restriction fragments. The assay was performed as described previously (11,12). Restriction fragments were filled in with [α-^{32}P]dATP. Labeled DNA (0.8 µg) was incubated with 50 ng of GST-fusion protein bound to glutathione–agarose beads for 1 h at 4°C in Lambda DNA-binding buffer [20 mM HEPES (pH 7.6), 1 mM EDTA (pH 8), 10 mM (NH$_4$)$_2$SO$_4$, 0.2% Tween-20, 1 mM DTT, 25 µg/ml BSA and 25 µg/ml poly(dI–dC)] plus varying amounts of KCl, as indicated in the text. The beads were washed three times with Lambda DNA-binding buffer minus DTT, BSA and poly(dI–dC). Bound DNA was eluted by boiling in Formamide loading buffer (90% formamide, 1× TBE, 0.04% bromophenol blue and 0.04% xylene cyanol), separated on a 6% sequencing gel and visualized by autoradiography.

For the oligonucleotide competition assays, 10 ng of ^{32}P-end-labeled double-stranded oligonucleotide was incubated with 100 ng of GST-fusion protein bound to glutathione beads in the Lambda DNA-binding buffer containing 50 mM KCl, 100 µg of salmon sperm DNA and varying amounts of unlabeled double-stranded competitor oligonucleotide, as indicated in the text. The beads were washed and the bound DNA was eluted and visualized as described above.

In vitro translation and DNA cellulose chromatography

The wild-type and mutant plasmid constructs were used to generate ^{35}S-methionine-labeled polypeptides using the TNT-coupled reticulocyte system (Promega). In vitro translated proteins were diluted in one bed volume (0.5 ml) of Column loading buffer [10 mM potassium phosphate (pH 6.2), 0.5% NP40, 10% glycerol, 1 mM DTT, aprotinin (1 mg/ml), pepstatin (1 mg/ml), leupeptin (1 mg/ml)], and applied to native DNA cellulose columns (Pharmacia). The protein sample was passed through the column twice. The columns used were Poly-Prep Chromatography Columns (Bio Rad catalog number 731-1550). Unbound material is

designated flow-through (FT). The columns were then washed multiple times with 1.0 bed volume column-loading buffer containing 50 mM NaCl (these are the 50 mM wash fractions), and eluted stepwise with column-loading buffer adjusted to contain increasing concentrations of NaCl from 100 to 800 mM, as indicated in the text. Fractions were analyzed by SDS–PAGE. The signal on the dried gel was quantified using a phosphorimager (Fuji) and associated software.

RESULTS

ARID subfamilies vary in sequence specificity and DNA-binding affinity

Human and mouse ARID-containing proteins can be classified into seven subfamilies: ARID1, ARID2, ARID3, ARID4, ARID5, JARID1 and JARID2. Within each designated subfamily, the degree of identity within the ARID regions is very high, ranging from 70 to 83% (Figure 1). In contrast, identity between ARID regions across subfamilies is <30%. Members within subfamilies generally also show clear relationships outside the ARID, as shown in Figure 2. These subclassifications are the basis for the current nomenclature of the ARID family, which has recently been accepted by the HUGO Gene Nomenclature Committee (HGNC) and the Mouse Genomic Nomenclature Committee (MGNC).

The DNA-binding properties of only a few ARID family members have been reported. Drosophila Dri and its murine ortholog Bright (ARID3A), as well as human MRF2 (ARID5B) all bind AT-rich sites selectively (3,4,6). However, human p270 (ARID1A), the closely related human protein ARID1B, and their apparent Drosophila and yeast counterparts, Osa and SWI1, all bind DNA without sequence specificity (9,11,12,20). A better understanding of the biological role of the ARID family will require a more thorough understanding of the distribution of sequence-specific DNA-binding properties among the individual members. We therefore undertook a survey designed to determine the DNA-binding properties of at least one member of each ARID subfamily.

Because amino acid identity within the ARID consensus is so high within subfamilies, originally a single member of each subfamily was selected to test for sequence specificity. Recombinant GST-fusion proteins were constructed using sequences that include the ARID domain of each protein examined. The sequence specificity of each protein was then examined in a DNA pull-down assay. This assay allows each protein access to a pool of Lambda DNA restriction fragments of varying size and sequence. As shown in Figure 3, Dri (the Drosophila counterpart of ARID3A) and MRF2 (ARID5B) bind in a sequence-specific manner in this assay, selectively binding to some fragments and not others. Selectivity for specific fragments becomes more pronounced in more stringent binding conditions (i.e. increased salt concentrations). Slight differences in the selected fragments between Dri and MRF2 probably reflect the fact that the two proteins select slightly different consensus sites in vitro (3,4,6). The major bands consistently selected by Dri in this assay are indicated by markers to the right of the Dri panel in Figure 3. In contrast to Dri, p270 (ARID1A) binds in a
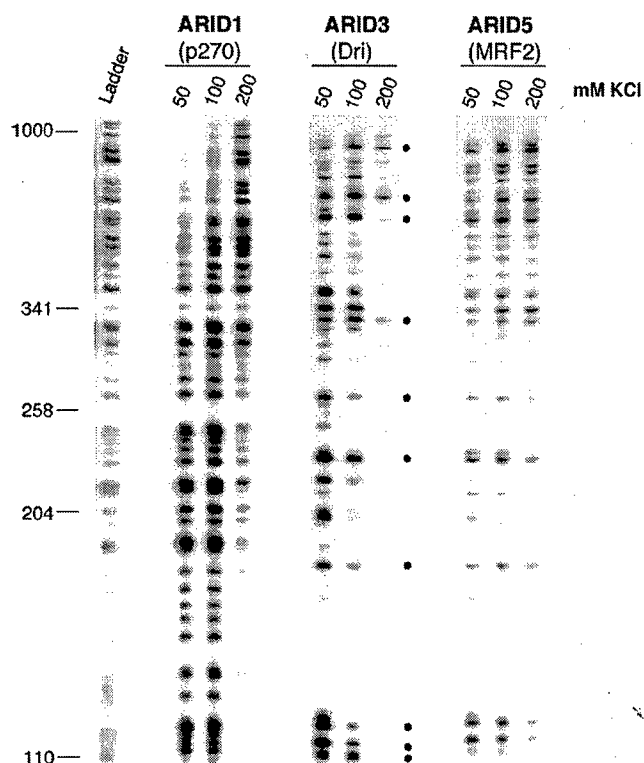
Figure 3. DNA-binding properties of the ARID family. Lambda phage DNA was digested with EcoRI, HindIII and Sau3AI to generate a large DNA oligonucleotide pool predicted to contain 128 fragments ranging in size from 12 to 2225 bp. The fragments were filled in with [^{32}P]dATP, incubated with GST-fusion proteins containing the ARID regions of each representative subfamily member as indicated, pulled down with glutathione beads, and analyzed by PAGE. Lane 1 shows the unselected pool of DNA fragments. Remaining lanes show the fragments selected in Lambda DNA-binding buffer with increasing KCl concentrations as indicated for p270, Dri and MRF2. Each subfamily is indicated at the top with the particular representative subfamily member assayed indicated directly below. The dots on the right of the Dri panel designate the major bands that are consistently selected by Dri in at least 10 repeats of this assay.

non-specific manner, binding to all fragments offered to it, showing selectivity only for longer fragments (>200 bp) at higher salt concentrations, presumably because longer fragments offer multiple binding sites. Despite the differences in sequence specificity, all three ARID proteins show similar affinities for DNA. These patterns have been documented previously (4,6,9,11), and are shown here for ease of comparison and as controls for the assay. The ARID1B protein has also been compared directly with p270, and found to bind without specificity (20).

The assay was used to examine the sequence specificity and DNA-binding affinity of representative members from each of the four remaining ARID subfamilies (Figure 4). ARID2 is the only member of its subfamily. A full-length human cDNA has not been reported thus far, but Genbank sequences predict an ARID consensus sequence at the N-terminus of the *ARID2* gene product. Isolation of N-terminal cDNA sequence by RT–PCR from the human liver cell line HepG2 confirms the presence of the ARID in the transcript (accession number AY727870.1). Studies

on mammalian ARID2 have not yet been reported, but the protein is an apparent ortholog of the *Drosophila* ARID protein BCDNA:GH12174 (CG3274). Both proteins contain an RFX domain, which is an additional DNA-binding domain [reviewed in (21)]. Interestingly, the protein product of *Drosophila* BCDNA:GH12174 was recently found to be a component of the SWI/SNF-like complex PBAP, and was designated BAP170 (22). This complex is distinguished from the BAP SWI/SNF-like complex in part by its lack of Osa. This finding extends the role of ARID-containing subunits as components of SWI/SNF-related chromatin-remodeling complexes. Analysis of ARID2 in the DNA pull-down assay (Figure 4) indicates that it binds DNA without sequence specificity, like all other known ARID-containing components of SWI/SNF-related complexes.

ARID4 subfamily DNA-binding activity is represented here by RBP1 (ARID4A). Amino acid identity within the ARID consensus is 75% between RBP1 (ARID4A) and RBP1L1 (ARID4B), the only other member of this class. The assay shown in Figure 4 indicates that RBP1 also binds DNA without sequence specificity. RBP1 has been characterized as a repressor of E2F-dependent transcription recruited by the retinoblastoma protein (pRb) and can recruit histone deacetylase (19,23,24). RBP1L1 (syn.: SAP180) is also able to repress transcription, at least when tethered to DNA through the Gal DNA-binding domain (25). Both RBP1 and RBP1L1/SAP180 have been found in association with the mSIN3-histone deacetylase complex (19,25).

JARID1 is the largest ARID subfamily. It contains four highly homologous members. RBP2 (JARID1A) can enhance nuclear hormone receptor transactivation in reporter assays (26). PLU-1 is highly expressed in breast cancers, and in reporter assays has transcriptional repressor properties (27). SMCX (JARID1C) and SMCY (JARID1D) are thought to be regulators of minor histocompatability antigen (28,29). The four JARID1 proteins share 83% amino acid identity within the ARID and are highly related across their full sequences. This subfamily, in common with JARID2, contains highly conserved JmJN and JmJC domains. The proposed nomenclature reflects these relationships. The function of JmJN and JmJC domains is not yet clear, but they exist in proteins other than ARID family members (30,31). Two representatives of the JARID1 subfamily were chosen for analysis. A second subfamily member was included for two reasons. First, amino acid identity between PLU-1 (JARID1B) and the other three members of this subfamily varies more than is typical within subfamilies in the Loop 2 and Helix 5 region of the ARID sequence (see Figure 1). PLU-1 has a histidine within the Helix 5 region at a position where the other members of this subfamily contain a leucine. This region is the major groove interaction site in other ARID members and could be expected to play an important role in sequence recognition. Second, PLU-1 is expressed in a highly tissue restricted manner in contrast to other JARID1 members, which are broadly expressed [reviewed in (1)]. RBP2 (JARID1A) was chosen to represent the more typical members of this subfamily. As shown in Figure 4, both RBP2 (JARID1A) and PLU-1 (JARID1B) bind DNA with little or no discernible sequence specificity.

The panel was completed by testing jumonji (JARID2), the only member of its subfamily. jumonji is developmentally
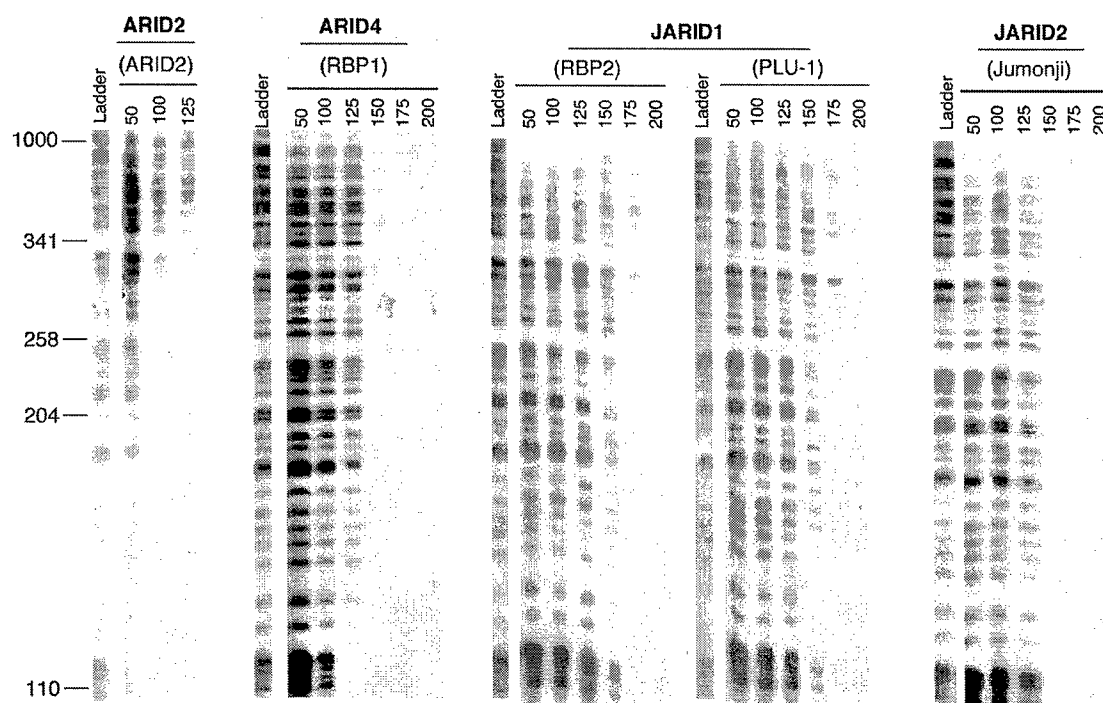
Figure 4. DNA-binding properties of the ARID family. Representative members of the remaining subfamilies were assayed as indicated in the legend to Figure 3. The unselected pool of DNA fragments is labeled Ladder and is shown for each individual experiment.

important in diverse organs (32,33), and can act as a transcriptional repressor in a reporter assay (34). Although jumonji has JmJN and JmJC domains in common with the JARID1 subfamily, the members of JARID1 are more similar to each other than to jumonji. Within the ARID domain, jumonji is only about 25% identical to members of the JARID1 group, which are 83% identical to each other. The jumonji ARID domain binds DNA in the pull-down assay without detectable sequence specificity (Figure 4). jumonji does show more of a tendency than other ARID family members to retain binding to lower molecular weight (<200 bp) DNA fragments even at high stringency, suggesting it does not disassociate as rapidly from DNA. This survey indicates that five of the seven ARID subfamilies bind DNA with no obvious sequence specificity. These results are summarized in Table 1.

The domains in the ARID1, ARID3 and ARID5 subfamilies retain DNA-binding affinity up to at least 200 mM KCl concentration (Figure 3). DNA affinity columns likewise indicate that p270 and Dri have similar DNA-binding affinities [(11) and Figure 8]. ARID1B, which is closely related overall to p270, retains DNA binding up to about 175 mM KCl [(20) and additional data not shown]. JARID1 subfamily domains also retain binding to at least 175 mM KCl (Figure 4). However, the data in Figure 4 indicate that ARID domains of the ARID2, ARID4 and JARID2 subfamilies have relatively low DNA-binding affinity. While this assay is not a direct measure of affinity, the results suggest that there are three distinguishable categories in the ARID family with regard to DNA-binding: sequence non-specific with low affinity, sequence non-specific with high affinity and sequence specific with high affinity. Previously, we showed that *Saccharomyces cerevisieae* SWI1 has relatively low affinity DNA-binding behavior that correlates with atypical

Table 1. Categorization of ARID subfamilies according to sequence specificity

HUGO nomenclature	Aliases	Tissue specificity
AT-specific		
ARID3A	Bright, DRIL1, E2FBP1	Restricted (mature B cells and testes) (3)
ARID3B	Bdp, DRIL2	Broad (52)
ARID3C	XM_071061	Not reported
ARID5A	MRF-1	Not reported
ARID5B	MRF2	Broad w/some specialization [high in brain, kidney, lung (39)]
Sequence non-specific		
ARID1A	p270, BAF250a, hOsa1, OSA1, B120, hSWI1, p250, SMARCF1	Broad (9,41,42,53,54)
ARID1B	pKIAA1235, BAF250b, p250R, hOsa2, held/OSA1	Broad (41,42,55,56)
ARID2	pKIAA1557	Broad[a]
ARID4A	RBP1	Broad (57)
ARID4B	RBP1L1, BCAA1, SAP180	Restricted (testes) (58)
JARID1A	RBP2	Broad (57)
JARID1B	PLU-1	Restricted (testes) (59)
JARID1C	SMCX, XE169	Broad (28)
JARID1D	SMCY, KIAA0234	Broad (28)
JARID2	jumonji	Specialized [brain, heart, skeletal muscles, kidney, thymus (33)]

[a]http://www.kazusa.or.jp/huge/gfpage/KIAA1557

sequence in the Loop 2 and Helix 5 region (11). Current results indicate that DNA-binding affinity of ARID family members can be low for reasons not easily apparent from inspection of the ARID sequence.
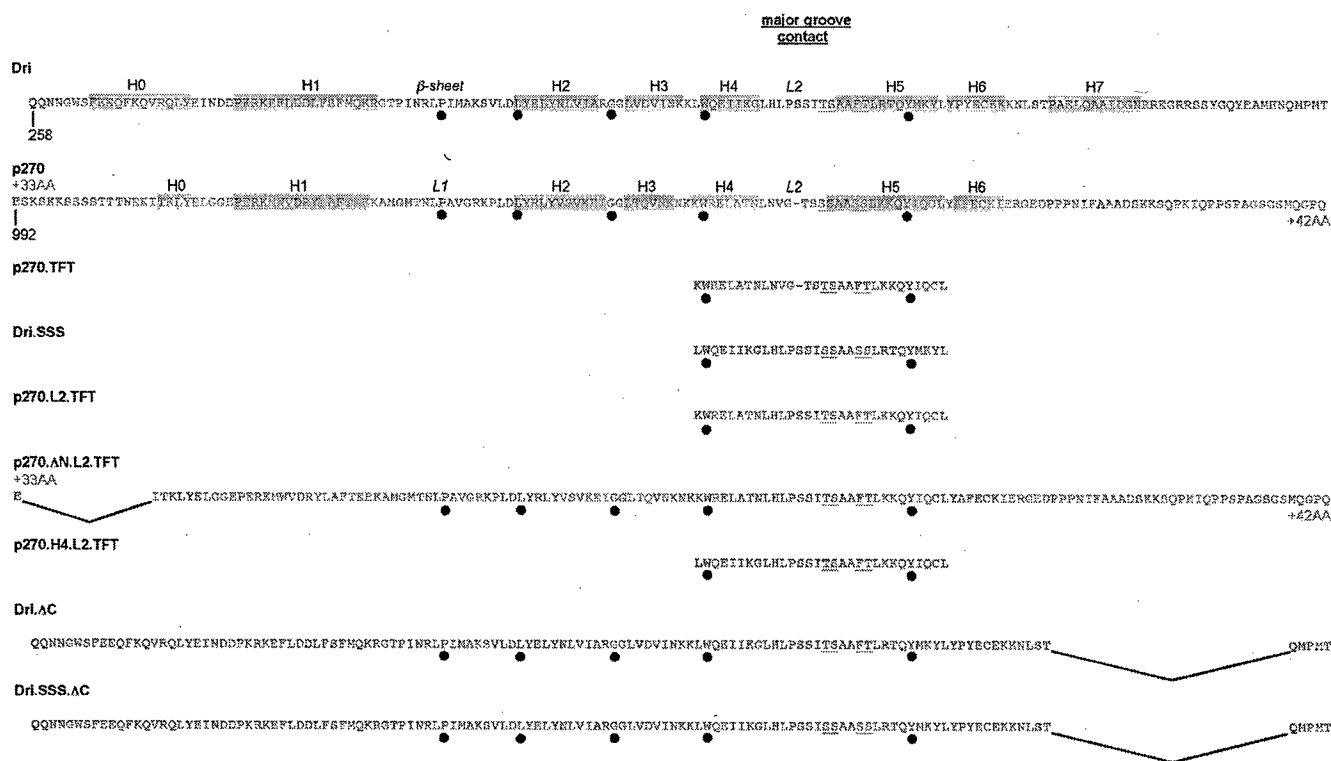
major groove
contact

Dri
HO   H1   β-sheet   H2   H3   H4   L2   H5   H6   H7

QQNNGWSFEEQFKQVRQLYEINDD...GTP INRLPIMAKSVLD...IARGG...VDVINKKL...KGLHLPSSITS...RTQYMKYL.YPYEC...KNL.ST...QMA...DGNRRRGRRSSYGQYEAMHNQMPHT
|
258

p270
+33AA   HO   H1   L1   H2   H3   H4   L2   H5   H6

ESKSKKSSSSTTTNSK...LYELGG...KAMGMTNLPAVGRKPLDLYRL...GG...TQVSKNKKW...LATNLNVG-TSS...QV...LY...EROEDPPPN.IFAAADSKKSQPKIQPPSPAGSGSMQGPQ
| +42AA
992

p270.TFT
                                        KWRELATNLNVG-TSTSAAFTLKKQYIQCL

Dri.SSS
                                        LWQEIIKGLHLPSSIGSAASSLRTQYMKYL

p270.L2.TFT
                                        KWRELATNLHLPSSITSAAFTLKKQYIQCL

p270.ΔN.L2.TFT
+33AA
E_____ITKLYELGGEPERKWVDRYLAFTEEKAMGMTNLPAVGRKPLDLYRLYVSVKE...GG.TQVSKNRKWRELATNLHLPSSITSAAFTLKKQYIQCLYAFECKIERGEDPPPNIFAAADSKKSQPKIQPPSPAGSGSMQGPQ
                                                                                                                    +42AA

p270.H4.L2.TFT
                                        LWQEIIKGLHLPSSITSAAFTLKKQYIQCL

Dri.ΔC
QQNNGWSFEEQFKQVRQLYEINDDPKRKEFLDDLFSFHQKRGTP INRLPIMAKSVLDLYELYNLVIARGGLVDVINKKLWQEIIKGLHLPSSITSAAFTLRTQYMKYLYPYBCEKKNLST_____QMPMT

Dri.SSS.ΔC
QQNNGWSFEEQFKQVRQLYEINDDPKRKEFLDDLFSFHQKRGTP INRLPIMAKSVLDLYELYNLVIARGGLVDVINKKLWQEIIKGLHLPSSIGSAASSLRTQYMKYLYPYRCEKKNLST_____QMPMT

Figure 5. Sequences of the p270 and Dri mutants. The amino acid sequences of the wild-type p270 and Dri ARIDs are aligned. The Dri sequence is shown in blue print. The p270 sequence is shown in red print, with identity to Dri shown in blue. The sequence shown for Dri is the expression product of plasmid p410. The p270 peptide used in these assays is the expression product of plasmid pNDB8, which is longer on each end than the Dri peptide. The number of additional amino acids on each side is indicated in the figure. For each sequence, the first residue is given a number corresponding to its position in the full-length protein (accession numbers p270: NP_006006, Dri: AAB05771). The five residues that are invariant among all known ARID sequences are indicated by dots. The α-helices determined from NMR studies are indicated by grey shading and are numbered (from H0 to H7) above each sequence, along with the loops (L1 and L2) and β-sheet. Both Dri and p270 ARIDs have been studied by NMR in complex with DNA (15,18).The Helix4–Loop2–Helix5 region is the helix–turn–helix motif that contacts the major groove in both proteins, although in p270 Loop 2 contacts seem to contribute less than in Dri. The proteins also contact the adjacent minor grooves. In Dri, this happens through the β-sheet and the end of Helix 7. p270 contacts the minor groove via the Loop 1 region that corresponds to the β-sheet, but the C-terminal area of the p270 ARID does not seem to contribute to DNA binding as much as this region does in Dri. Additionally, a region of about 15 amino acids upstream of p270 Helix 0 interacts with DNA, but a comparable contact site does not exist in Dri. According to the structural study of the Dri–DNA complex, four residues in the Loop2–Helix5 region, two threonines (T), one serine (S) and one phenylalanine (F), make base-specific contacts. These residues and the corresponding residues in p270, all serines (S), are indicated by underlining in the figure. The mutant peptides p270.TFT, Dri.SSS, p270.L2.TFT and p270.H4.L2.TFT have changes only in the Helix4–Loop2–Helix5 region and therefore only that region is shown. For the in-frame deletion mutants, the whole sequence is shown, with the boundaries between deleted sequences shown by the solid lines. For the p270.ΔN.L2.TFT mutant, residues from S993 to K1007 were deleted. For the Dri.ΔC and the Dri.SSS.ΔC mutants, residues from P378 to N405 were deleted.

Sequence specificity does not depend solely on the identity of the specific major groove contact residues of Dri

Inspection of the sequences in Figure 1 does not reveal any obvious distinction between sequence-specific and sequence-non-specific ARIDs. However, the structures of the MRF2, Dri and p270 ARIDs have been solved in complex with DNA (15,17,18). Each study agrees that a portion of the region encompassing Loop 2 and/or Helix 5 lies within the major groove (see Figure 5), and that regions upstream and/or downstream of the junction of Loop 2 and Helix 5 contact the minor groove. The results have generated some ideas about the basis for sequence specificity, but these have not yet been tested empirically. Iwahara *et al.* (18) studied the Dri ARID by NMR, and identified four residues in Loop 2 and Helix 5 of Dri that make base-specific interactions in the major groove of the DNA. These residues included two threonines (T), a serine (S) and a phenylalanine (F), and are underlined in the Dri

ARID sequence shown in Figure 5. p270 has a serine (S) at each of the corresponding positions. This study suggested that the lack of the threonines and of a non-polar residue at the phenylalanine position underlies the lack of sequence specificity in p270.

We undertook a site-specific mutagenesis study to compare the role of individual elements of the ARID in the determination of sequence specificity. The structures of the Dri and p270 ARIDs are the best characterized among their respective types in regard to DNA interactions, so these domains were chosen for comparison. To test whether the presence of the threonine and phenylalanine residues in Helix 5 is sufficient to confer sequence specificity, we generated a mutant variant in which threonines and a phenylalanine were introduced into the appropriate positions in p270. The sequence of the resultant mutant, designated p270.TFT, is shown in Figure 5.

The behavior of the p270.TFT mutant was examined in the Lambda DNA pull-down assay. The results (Figure 6A) show
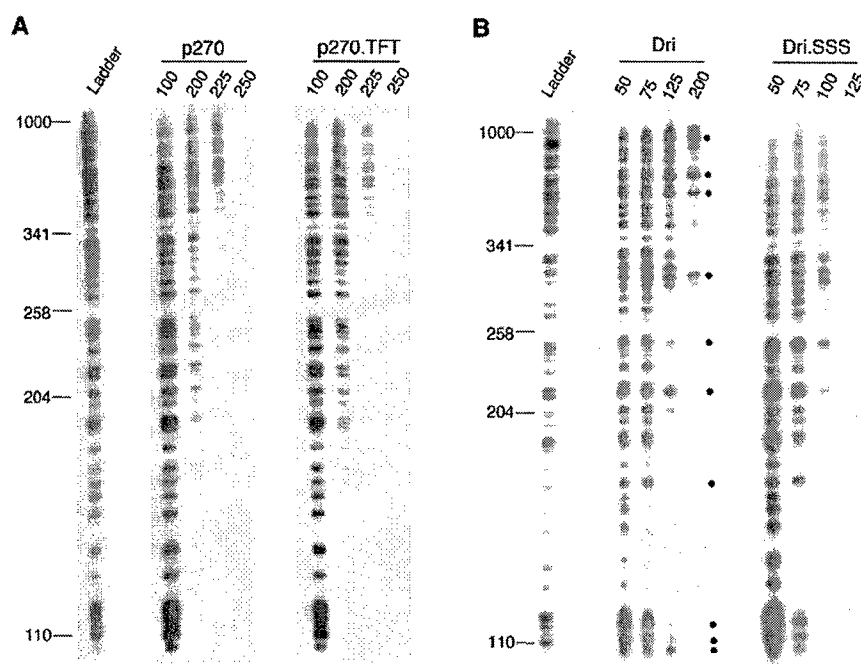
Figure 6. Assay of substitutions in the Helix 5 contact residues. The sequence specificity of the p270.TFT (**A**) and Dri.SSS (**B**) mutant peptides was tested in the Lambda DNA pull-down assay as described in the legend to Figure 3. The profiles of the wild-type p270 and Dri peptides are shown for reference. The dots on the right of the Dri panel designate the major bands that are consistently selected by Dri in this assay.

the mutant behaves exactly like wild-type p270. The substitutions do not confer any detectable capacity for sequence-specific binding, even at the highest stringency.

We also generated the reverse substitution in the Dri construct, replacing the presumptive base-specific contact residues with serines. The sequence of the resultant mutant, designated Dri.SSS, as shown in Figure 5, and the DNA-binding behavior is shown in Figure 6B. Strikingly, the Dri.SSS variant maintains a clear capacity for sequence-specific binding, selecting a pattern of DNA fragments very similar to those selected by wild-type Dri. It is apparent, though, from the KCl titration in Figure 6B, that the Dri.SSS variant has reduced overall affinity for DNA. No DNA binding is detected at this exposure in the 125 mM lane, while wild-type Dri consistently shows detectable binding in similar assays to at least 200 mM KCl (Figures 3 and 6B). The most direct explanation for these results is that these positions in Dri do make significant DNA contacts that are important for affinity, but which are not major determinants of sequence specificity. The DNA-binding affinity of p270 is strong despite the presence of serines at these positions, implying that the role of individual positions is not neccesarily directly comparable between different ARIDs.

The role of the helix–turn–helix motif

The NMR-derived p270 structure was reported earlier this year (15) and compared directly with the MRF2 structure (17). The structure indicates that Helix 5 of p270 lies within the major groove. However, these authors determined, by assessment of changes in the dynamics of the complex, that the shorter Loop 2 of p270 is less flexible than the corresponding loop in MRF2.

This suggested a 'folding upon binding' mechanism of sequence recognition, in which the shorter length and/or less flexible composition of Loop 2 of p270 in comparison to MRF2 and Dri affects orientation of the major groove contact residues, and is thus responsible for lack of sequence-specific contact.

p270 Loop 2 does not appear to contact DNA directly (15), but to evaluate the possibility that Loop 2 affects the orientation of Helix 5 within the major groove, a p270 chimera was constructed in which the Loop 2 sequence of Dri was placed in the p270.TFT construct. This chimera, designated p270.L2.TFT, contains the major groove contact residues of Dri as well as a Loop 2 sequence derived entirely from Dri, which, therefore, should be sufficiently long and flexible to permit proper orientation of the DNA contact residues within the major groove. Nonetheless, when tested in the DNA pull-down assay, the p270.L2.TFT chimera shows no greater tendency to sequence specificity than wild-type p270, and, indeed, has slightly less overall affinity for DNA (Figure 7). Since the TFT substitution alone did not affect p270 DNA-binding affinity, it is likely that the introduction of the exogenous Loop 2 sequence created a distortion that interferes with the overall strength of DNA contact in the p270 ARID.

A significant difference between the way the p270 and Dri ARIDs interact with DNA is that the p270 ARID has an additional large minor groove interaction site just upstream of Helix 0 (15). We considered the possibility that this region of 15 amino acids interacts strongly and non-specifically with DNA in a way that masks potential sequence-specificity in p270. We therefore deleted this segment in the p270.L2.TFT chimera to generate a new construct designated p270. ΔL2.TFT. The DNA-binding affinity of this fragment is
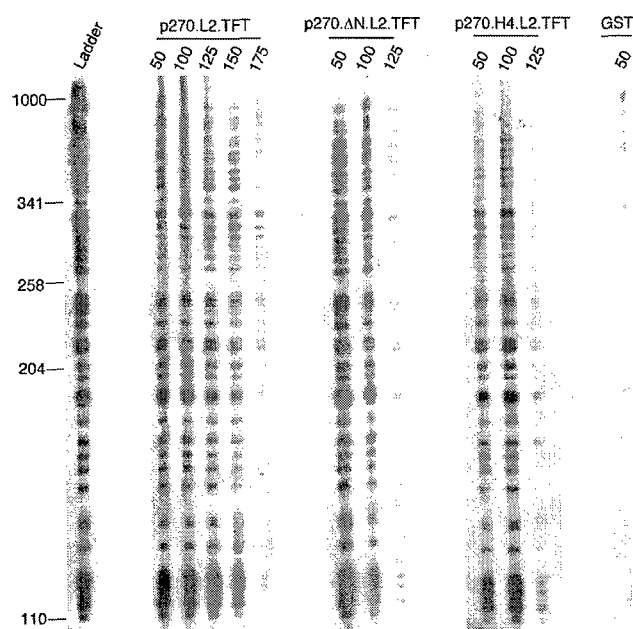
Figure 7. Assay of p270/Dri chimeras. The sequence specificity of the p270.L2.TFT, p270.ΔN.L2.TFT and p270.H4.L2.TFT mutant peptides was tested with the Lambda DNA pull-down assay as described in the legend to Figure 3. The GST DNA-binding profile is shown as a control.
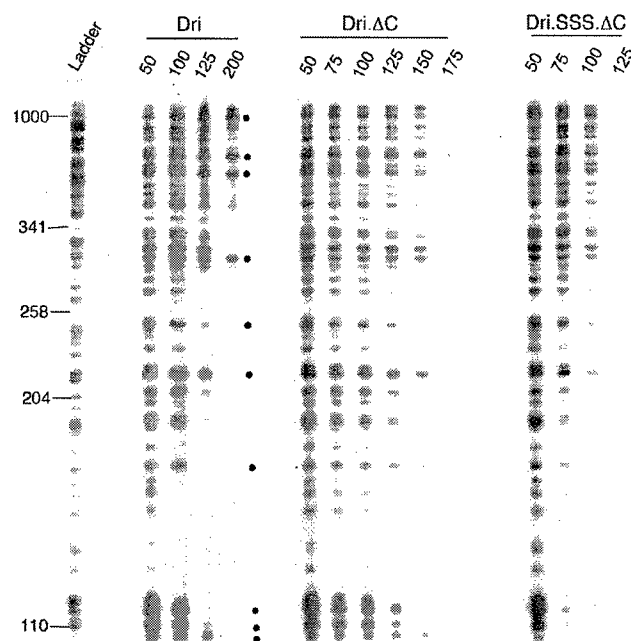
Figure 8. Deletion of the ARID3 extended sequence. The sequence specificity of the Dri.ΔC and Dri.SSS.ΔC mutant peptides was tested with the Lambda DNA pull-down assay as described in the legend to Figure 3. The profile of the wild-type Dri, along with the markers for the major bands selected by Dri, are also shown for reference.

reduced still further, confirming that the N-terminal region contributes significantly to DNA contact. However, the peptide still shows little or no selection for specific fragments (Figure 7). This argues against the possibility that sequence selectivity was transferred by the introduction of the Loop 2 and Helix 5 residues of Dri, but was masked by the unique N-terminal contact region of p270.

The ARID is categorized as a modified helix–turn–helix motif-based DNA-binding domain, in which the second helix of the motif (Helix 5) is the recognition helix. To test the possibility that the first helix of the motif (Helix 4) influences the orientation of Loop 2 and the recognition helix, the p270.L2.TFT construct was further modified such that the entire region from the beginning of Helix 4 to the last major groove contact residue consists of contiguous Dri sequence. The name of this mutant is p270.H4.L2.TFT. This construct still binds to DNA without any clear sequence selectivity (Figure 7). Moreover, affinity is reduced below that of the p270.L2.TFT variant, supporting the suggestion above, that introduction of exogenous sequence creates distortions that interfere with the overall strength of DNA contact in the p270 ARID. Individual elements are not directly exchangeable between different ARIDs. Together, these results indicate that sequence specificity in the ARID does not depend solely on the specific amino acid composition in the major groove contact region.

Contribution of the extended ARID region

Members of the ARID3 subfamily in all species studied are characterized by the presence of an 'extended' ARID sequence, a region of very high identity (>70% identity across ∼35 residues) immediately following the core ARID

consensus (see Figure 1). This region includes Helix 7, which is so far unique to the ARID3 subfamily, and extends beyond it. The extended ARID region has been identified as a DNA contact region in Dri (18). The extended ARID sequence is not present in the ARID5 subfamily, so cannot be a required determinant of sequence specificity. However, a corresponding position C-terminal to the core ARID consensus in MRF2 has been identified as a DNA contact region (17). In contrast, the corresponding region in p270 does not appear to make significant DNA contact (15). To assess the contribution of this region of Dri to sequence specificity, an in-frame deletion of sequence encoding 28 amino acids was generated in this region. The resulting mutant is designated Dri.ΔC. The DNA pull-down assay indicates that this construct retains a considerable measure of sequence selectivity (Figure 8). The construct shows slightly reduced DNA-binding affinity, consistent with the interpretation that the deleted region includes a DNA contact site.

The deletion of the extended ARID was also engineered in the Dri.SSS background. The resulting mutant is designated Dri.SSS.ΔC. In the DNA pull-down assay, the Dri.SSS.ΔC construct (Figure 8) has the same reduced DNA-binding affinity as was seen in the Dri.SSS mutant in Figure 6B. The sequence selectivity is further reduced, but a weak selectivity is still evident.

The lambda DNA restriction fragment pool was used as the target DNA in order to offer a wide range of sequence possibilities to the ARID proteins used in this survey. This allowed for the possibility that some family members, or some mutant variants, would show sequence preference, but for a previously unrecognized sequence. However, a disadvantage
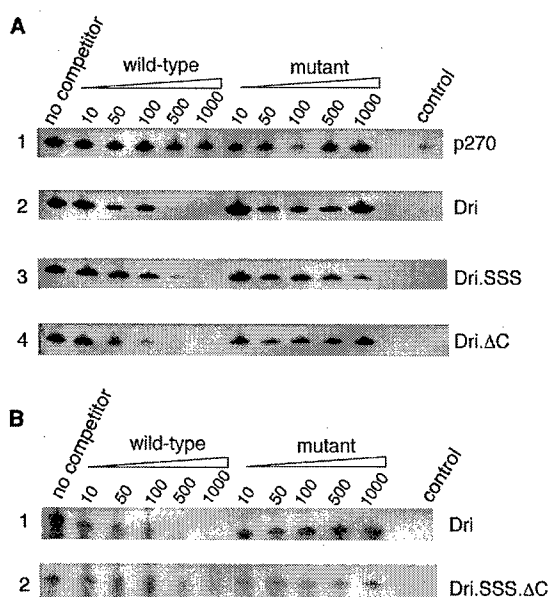
Figure 9. Competition assay with the Dri mutants. The competition assays show binding of the GST-fusion peptides to a labeled oligonucleotide containing a consensus sequence for Dri. All reactions contain a 10 000-fold excess of salmon sperm DNA. Binding to the consensus sequence was competed with increasing amounts of unlabeled specific competitor, either the consensus sequence (wild-type) or an altered sequence (mutant) in a 10-fold, 50-fold, 100-fold, 500-fold or 1000-fold excess. A separate reaction was used in each experiment with just glutathione–agarose beads as a control.

of the lambda DNA restriction fragment pool is that the complex restriction pattern precludes the identification of individual restriction fragments, or the actual sequence of the selected fragments. To obtain a more quantitative measurement of sequence specificity, selected mutants were probed in an oligonucleotide competition assay, where their affinity for a Dri consensus binding site (CCAATTAATCCC) was compared with their affinity for an altered consensus site (CCAATTGCTCCC). The consensus sites were synthesized as three tandem repeats. This assay was performed in low salt (50 mM) conditions, so that the effect of increasing salt concentrations on the conformation of the protein would not be a factor in the assay. The results are shown in Figure 9A. The Dri peptide shows a clear preference for its identified consensus site in this assay, as reported previously (4). A 500-fold excess of cold competitor with the correct consensus sequence competes effectively with the labeled probe, while the altered sequence, even at 1000-fold excess, shows little ability to displace the peptide (Figure 9A, panel 2). In contrast, the AT-rich consensus site does not compete for p270 binding any better than the mutant oligonucleotide (Figure 9A, panel 1).

The Dri.SSS and Dri.ΔC variants were both tested in this assay. The results show that each has a higher preference for the AT-rich consensus site than for the altered oligonucleotide (Figure 9A, panels 3 and 4), meaning that they clearly retain sequence-specific binding. This is consistent with the behavior they showed in Figures 6B and 8.

When we attempted this assay with the Dri.SSS.ΔC construct, we found that the peptide bound poorly to the oligonucleotide even at 50 mM salt. Because the ARID proteins show a generally higher affinity for longer pieces of DNA, we

attempted the assay with a longer oligonucleotide, containing eight repeats of the consensus sequence rather than three. The Dri.SSS.ΔC peptide bound well to this probe (Figure 9B). Wild-type Dri showed the same behavior on this probe as on the shorter one: it was displaced more readily by the true consensus sequence than by the altered sequence (Figure 9B, panel 1). The Dri.SSS.ΔC peptide showed less specificity than wild-type Dri, but a weak selectivity was still evident (Figure 9B, panel 2), consistent with the behavior seen in Figure 8.

The DNA-binding phenotype of the double mutant, Dri.SSS.ΔC, implies that the region C-terminal to the core ARID consensus, and amino acid identity at the major groove contact site, contribute to the presence of sequence specificity in ARID3 subfamily proteins, but do not support a conclusion that small amino acid differences, such as the identity of residues at the junction of Loop 2 and Helix 5, or the length of Loop 2, are the principal determinants of sequence specificity. Rather, the results suggest that overall differences in the three-dimensional structure of individual ARID subfamilies determines the presence of sequence specificity. A similar situation appears to hold for the distinction between sequence-specific and sequence-non-specific DNA binding in HMG domain proteins, considered further in the Discussion.

p270 and Dri differ in their ability to tolerate mutations in the aromatic scaffold

The potential for differences in the overall structure of the p270 and Dri ARIDs was probed by introducing changes into the aromatic scaffold of the two domains. Within the core consensus sequence, there are five invariant amino acids that are almost identically spaced across each ARID. These are indicated by red text in Figure 1 and dots in Figure 4, and include a tryptophan (W) in Helix 4, a tyrosine (Y) in Helix 5 and a proline (P) in Loop 1. The presence of a series of invariant aromatic residues has been recognized as a structural scaffold in other helix–turn–helix motifs, including the DNA-binding motif in c-Myb and the homeodomain (35,36).

To test the contribution of the invariant residues in the ARID structure, specific invariant residues were changed to the small neutral residue alanine in the ARIDs of both Dri and p270. The resultant wild-type and mutant peptides were translated *in vitro*, and their DNA-binding affinity was assessed using a sensitive and quantitative DNA affinity column chromatography assay described previously (9,11). Because the DNA is in large excess, the assay is unbiased with respect to sequence specificity. The results are shown in Figure 10.

The interaction of the wild-type p270 ARID-containing peptide with DNA is as strong as that of the wild-type Dri ARID-containing peptide. In both wild-type proteins, 80–90% of the signal is retained on the columns. The remainder comes off in the flow-through and the first wash, and presumably represents a fraction of peptide that did not bind due to impaired folding. The proline-to-alanine substitution has very little effect on the elution profile of either p270 or Dri, suggesting that this residue, though invariant, is not by itself critical for the maintenance of structural integrity in the domain.

On the other hand, the Helix 4 tryptophan-to-alanine substitution seriously impaired binding to native DNA in both
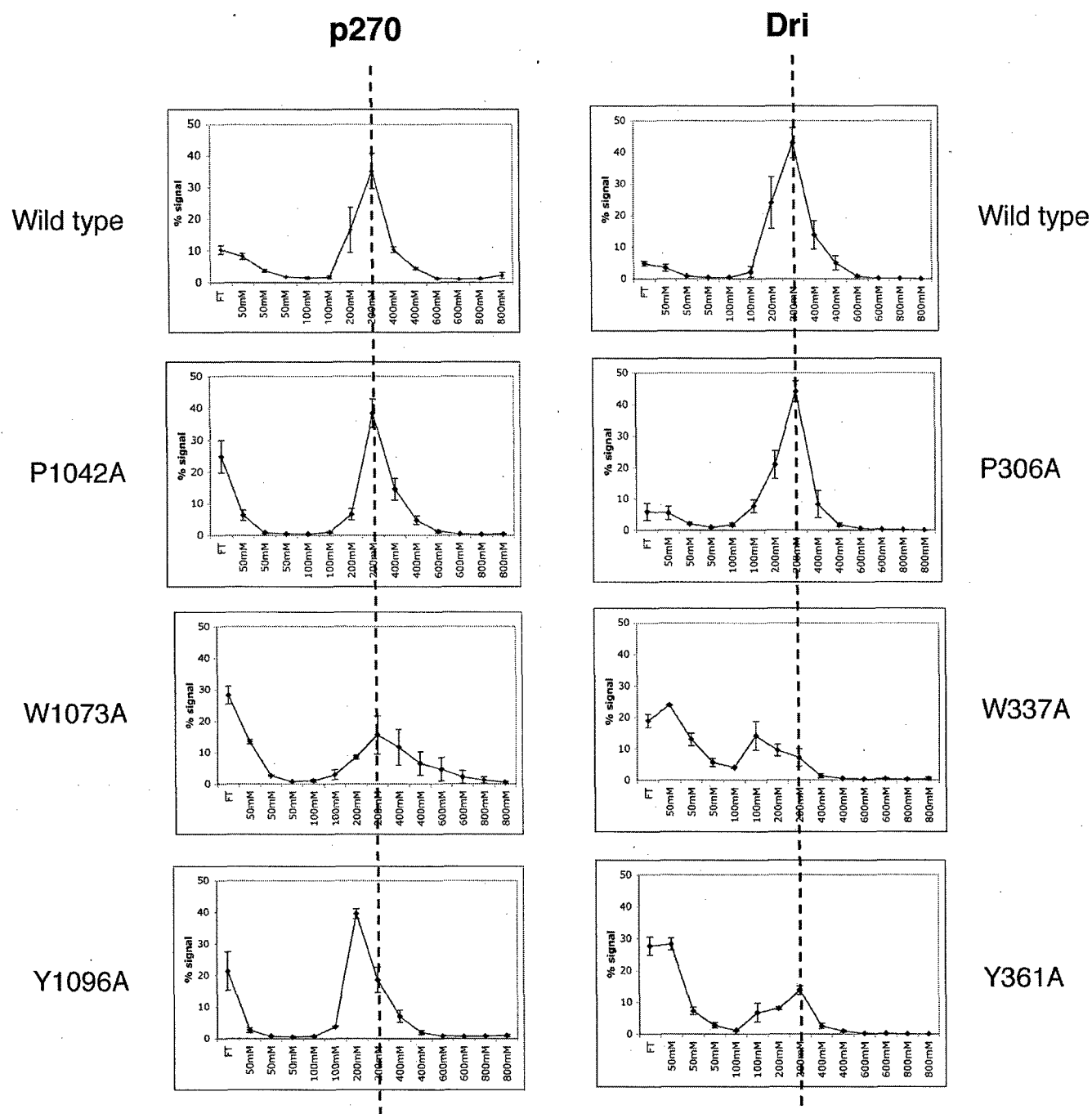
p270

Dri

Figure 10. Substitution of invariant residues. The strength of the interaction of the wild-type p270 and Dri peptides, as well as peptides where invariant residues were substituted by an alanine, was tested by DNA affinity chromatography. *In vitro* translated ^{35}S-methionine-labeled peptides were applied to a native DNA cellulose column as described in Materials and Methods. Bound protein was eluted stepwise with loading buffer adjusted to contain increasing concentrations of NaCl from 100 to 800 mM, as indicated in the figure. Fractions were separated by SDS–PAGE and the p270 signal in each fraction was quantified by phosphorimaging. The results are plotted as the percentage of signal in each fraction relative to the entire signal recovered. Each experiment was performed at least twice and the error bars represent the average deviation. Graphs are aligned for ease of comparison. The dashed line indicates the second 200 mM fraction for reference.

ARIDs. Much of the mutant protein (about 45–50% in either p270 or Dri) fails to bind to the column and is recovered in the flow-through and wash fractions. The strongly deleterious effect of the tryptophan substitution suggests that the invariable tryptophan plays a critical role in maintaining the overall integrity of the ARID structure in both sequence-specific and sequence-non-specific representatives of the family.

p270 and Dri showed a different tolerance to the third mutation, a tyrosine-to-alanine substitution in Helix-5. The elution profile of the p270 mutant peptide is similar to that

p270

Wild type

P1042-Y1096A
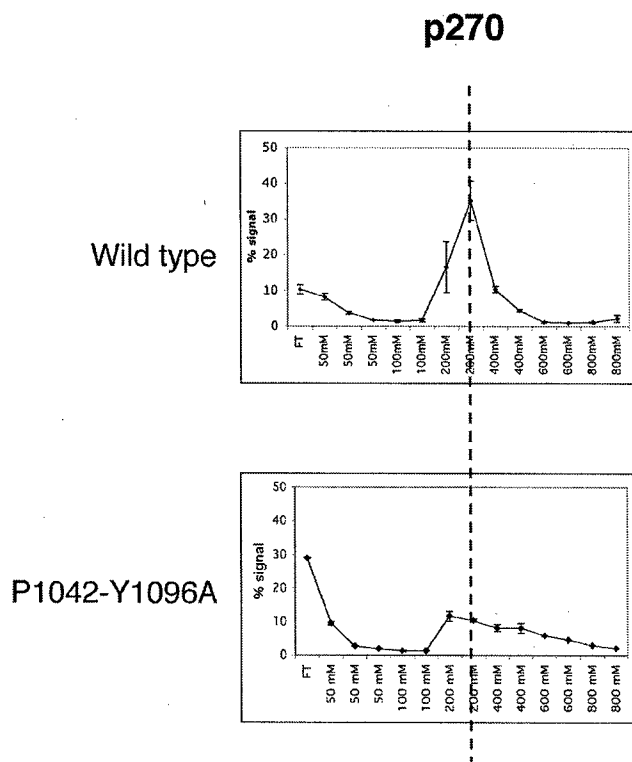
Figure 11. Combination of the proline and tyrosine substitutions can act synergistically to impair p270 ARID binding to DNA. The strength of the interaction of the combined substitution mutant was tested by DNA affinity chromatography as described above. The elution profile of the wild-type p270 is repeated in this panel and the graphs are aligned for ease of comparison. The dashed line indicates the second 200 mM fraction for reference.

of the wild-type peptide. Approximately the same amount of signal is retained on the column, although the shift in the elution peak from the second to the first 200 mM fraction indicates a weakening of affinity. This type of elution profile suggests that the substitution causes loss of one or more DNA contact sites, but does not suggest that protein folding is grossly affected. In contrast, the corresponding substitution in Dri is as deleterious as the tryptophan substitution, with 40–50% of the signal failing to bind to the column, implying that this residue is critical in the Dri ARID for maintaining proper structure.

To probe further the role of the Helix 5 tyrosine in the p270 ARID, the Y1096A substitution was combined with the P1042A substitution. The effect of the combined mutations was highly synergistic (Figure 11), generating a DNA-binding profile almost as defective as that seen with the W1073A substitution. This confirms that the Helix 5 tyrosine is important to structural integrity in the p270 ARID, but the results suggest that the p270 ARID is more able than the Dri ARID to tolerate changes in its aromatic scaffold. Thus, there appear to be fundamental differences in the ARID structures of p270 and Dri that go beyond simple differences at specific amino acid positions. This is consistent with the detrimental effects observed above of exchanging presumably analogous sequences between the two proteins. The mutagenesis studies argue against a conclusion that

specific amino acids in Loop 2 or Helix 5 are the main determinants of sequence specificity. Most probably, this is determined by multiple interacting differences across the entire ARID structure.

DISCUSSION

The overall conclusion from the survey described here is that the majority of ARID subfamily domains bind DNA without regard to sequence specificity. Thus, the acronym is somewhat of a misnomer, although it is a well established and useful descriptor for a domain whose parameters are well-defined. This survey did not probe the behavior of every single member of the human ARID family. The proteins that have not been tested directly, here or elsewhere, among the subfamilies now designated as sequence non-specific are RBP1L1 (ARID4B), SMCX and SMCY (JARID1C and JARID1D). Each of these shows at least 75% identity and even greater similarity to the tested members of its subfamily. In addition, a mention of data not shown in a report on RBP1L1 (syn:SAP180) notes that a high-affinity consensus binding site could not be found in DNA-binding site selection experiments (25). Among the subfamilies now designated as AT-specific, only DRIL2 (ARID3B) and ARID3C have not been tested empirically, but again, there is at least 75% identity and more than 90% similarity between these ARID sequences and the AT-specific prototypes Bright and Dri. There is a potential conflict between our conclusions and a report suggesting that an ARID-containing fusion peptide of jumonji (JARID2) may have general selectivity for AT-rich sequences, since a majority of sequences selected by jumonji from a pool of random oligonucleotides were AT rich (34). However, several sequences that jumonji bound with equally high affinity in that study were not AT rich, and a precise consensus site could not be identified. The present survey is concerned with the properties inherent in the ARID sequence from each subfamily. As such, it was conducted with fusion proteins expressing the respective ARID sequences separate from the context of the native proteins. It remains possible that the endogenous proteins acquire a degree of sequence-specific binding behavior in physiological conditions.

The emergence of specific ARID subfamilies appears to have occurred early in evolution. *S.cerevisieae* encodes two ARID proteins. The ARID sequences do not correlate closely with any particular human subfamily, but overall the proteins seem most similar to the ARID1 and JARID1 subfamilies. *Schizosaccharomyces pombe* encodes four ARID proteins, two that are members of chromatin remodeling complexes and two that share similarity to the JARID1 subfamily. *Ceanorhabditis elegans* encodes four ARID proteins, aligning with human subfamilies ARID1, ARID2, ARID3 and JARID1, thus including a single AT-specific subfamily representative (for an excellent review of ARID evolution see http://www.lifesci.utexas.edu/research/tuckerlab/bright/evolution/). The ARID protein CFI-1 is the only identified member of an ARID3-type subfamily within *C.elegans* and prefers the same AT-rich consensus sequence as Dri in a competition assay (37). *Drosophila melanogaster* encodes six ARID proteins, one aligning with each subfamily except the second AT-rich specific subfamily ARID5. These patterns suggest that

ARIDs probably began as sequence non-specific and gained the property of sequence specificity through evolution.

The precise function of all the human ARID proteins is not known. Members of the AT-specific ARID3 and ARID5 subfamilies are sequence-specific transcription factors with recognized promoter targeting functions and important roles in development and differentiation (3,4,5,38–40). Among the sequence-non-specific ARID proteins, several appear to participate in general transcription and chromatin remodeling functions. ARID1A and ARID1B are mutually alternative members of human SWI/SNF-related complexes (20,41,42) and ARID1A (p270) is implicated in the tumor suppressor activity of the complexes (43). Human ARID2 is uncharacterized, but the *Drosophila* ortholog of ARID2 is a member of a SWI/SNF-like complex (22). ARID4A and ARID4B can associate with the mSIN3-histone deacetylase complex (19,25). Members of the JARID1 and JARID2 subfamilies show transcription activation and/or repression functions (26,27,34). To date, only the Dri and Bright (ARID3A) ARIDs have actually been shown to be required for the physiological function of their cognate proteins (44,45). The ARID of the *S.cerevisieae* protein SWI1 appears dispensable for complementation of the SWI1 phenotype (46), but transient reporter assays suggest the ARID is required for a transactivation function in human ARID1B (41). More physiological experiments are needed.

Site-specific mutagenesis has not revealed any precise determinants for sequence specificity or lack of it within the ARID family. Most probably, this is determined by multiple interacting differences across the entire ARID structure. A similar situation appears to hold for the distinction between sequence-specific and sequence-non-specific DNA binding in high mobility group (HMG) domain proteins. HMG domain containing proteins bind DNA through contacts in the minor groove. They recognize DNA structures such as four-way junctions, distorted cisplatin-kinked DNA and supercoiled DNA, and generally have the ability to bend DNA. One HMG protein subfamily consists of transcription factors like LEF-1 (lymphoid enhancer factor-1) and SRY (mammalian sex determining gene) that bind sequence specifically to AT-rich sequences in enhancer and promoter regions. Members of this subfamily contain one copy of the HMG domain and are tissue specific. Another subfamily comprises chromosomal proteins such as HMG1 and HMG2 that bind DNA in a sequence-non-specific manner. These proteins generally contain two or more HMG domains (47,48). There is a high degree of sequence similarity and structural characteristics between the sequence-specific and the sequence-non-specific HMG domains in complex with DNA. Some highly conserved residues have been identified as very important in sequence specificity of HMG domains, but these residues alone are not the sole determinant of sequence specificity [reviewed in (47)]. Rather, sequence-specificity appears to be a combination of effects of residues on the domain's positioning, affinity, its stability in complex with DNA, the number of interactions on the protein–DNA interface and the number of base-specific contacts (49). Studies of the HMG domain indicate the difference between sequence-specific and sequence-non-specific members of the same family is generally more complex than the simple substitution of contact residues for neutral residues.

REFERENCES

1. Wilsker,D., Patsialou,A., Dallas,P.B. and Moran,E. (2002) ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development. *Cell Growth Differ.*, **13**, 95–106.
2. Kortschak,R.D., Tucker,P.W. and Saint,R. (2000) ARID proteins come in from the desert. *Trends Biochem. Sci.*, **25**, 294–299.
3. Herrscher,R.F., Kaplan,M.H., Lelsz,D.L., Das,C., Scheuermann,C. and Tucker,P.W. (1995) The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. *Genes Dev.*, **9**, 3067–3082.
4. Gregory,S.L., Kortschak,R.D., Kalionis,B. and Saint,R. (1996) Characterization of the dead ringer gene identifies a novel, highly conserved family of sequence-specific DNA binding proteins. *Mol. Cell. Biol.*, **16**, 792–799.
5. Huang,T.H., Oka,T., Asai,T., Okada,T., Merrills,B.W., Gertson,P.N., Whitson,R.H. and Itakura,K. (1996) Repression via differentiation-specific factor of the human cytomegalovirus enhancer. *Nucleic Acids Res.*, **24**, 1695–1701.
6. Whitson,R.H., Huang,T. and Itakura,K. (1999) The novel Mrf-2 DNA-binding domain recognizes a five-base core sequence through major and minor-groove contacts. *Biochem. Biophys. Res. Commun.*, **258**, 326–331.
7. Dallas,P.B., Yaciuk,P. and Moran,E. (1997) Characterization of monoclonal antibodies raised against p300: both p300 and CBP are present in intracellular TBP complexes. *J. Virol.*, **71**, 1726–1731.
8. Dallas,P.B., Cheney,I.W., Liao,D., Bowrin,V., Byam,W., Pacchione,S., Kobayashi,R., Yaciuk,P. and Moran,E. (1998) p300/CREB binding protein-related protein p270 is a component of mammalian SWI/SNF complexes. *Mol. Cell. Biol.*, **18**, 3596–3603.
9. Dallas,P.B., Pacchione,S., Wilsker,D., Bowrin,V., Kobayashi,R. and Moran,E. (2000) The human SWI/SNF complex protein, p270, is an ARID family member with nonsequence-specific DNA binding activity. *Mol. Cell. Biol.*, **20**, 3137–3146.
10. Nie,Z., Xue,Y., Yang,D., Zhou,S., Deroo,B.J., Archer,T.K. and Wang,W. (2000) A specificity and targeting subunit of a human SWI/SNF family-related chromatin-remodeling complex. *Mol. Cell. Biol.*, **20**, 8879–8888.
11. Wilsker,D., Patsialou,A., Zumbrun,S.D., Kim,S., Chen,Y., Dallas,P.B. and Moran,E. (2004) The DNA-binding properties of the ARID-containing subunits of yeast and mammalian SWI/SNF complexes. *Nucleic Acids Res.*, **32**, 1345–1353.
12. Collins,R.T., Furukawa,T., Tanese,N. and Treisman,J.E. (1999) Osa associates with the Brahma chromatin remodeling complex and promotes the activation of some target genes. *EMBO J.*, **18**, 7029–7040.
13. Yuan,Y.C., Whitson,R.H., Liu,Q., Itakura,K. and Chen,Y. (1998) A novel DNA-binding motif shares structural homology to DNA replication and repair nucleases and polymerases. *Nature Struct. Biol.*, **5**, 959–964.

14. Iwahara,J. and Clubb,R.T. (1999) Solution structure of the DNA binding domain from Dead ringer, a sequence-specific AT-rich interaction domain (ARID). *EMBO J.*, **18**, 6084–6094.

15. Kim,S., Zhang,Z., Upchurch,S., Isern,N. and Chen,Y. (2004) Structure and DNA-binding sites of the SWI1 AT-rich interaction domain (ARID) suggest determinants for sequence-specific DNA recognition. *J. Biol. Chem.*, **279**, 16670–16676.

16. Tu,X., Wu,J., Xu,Y. and Shi,Y. (2001) 1H, 13C and 15N resonance assignments and secondary structure of ADR6 DNA-binding domain. *J. Biomol. NMR*, **21**, 187–188.

17. Zhu,L., Hu,J., Lin,D., Whitson,R., Itakura,K. and Chen,Y. (2001) Dynamics of the Mrf-2 DNA binding domain free and in complex with DNA. *Biochemistry*, **40**, 9142–9150.

18. Iwahara,J., Iwahara,M., Daughdrill,G.W., Ford,J. and Clubb,R.T. (2002) The structure of the Dead ringer–DNA complex reveals how AT-rich interaction domains (ARIDs) recognize DNA. *EMBO J.*, **21**, 1197–1209.

19. Lai,A., Kennedy,B.K., Barbie,D.A., Bertos,N.R., Yang,X.J., Theberge,M.C., Tsai,S.C., Seto,E., Zhang,Y., Kuzmichev,A., Lane,W.S., Reinberg,D., Harlow,E. and Branton,P.E. (2001) RBP1 recruits the mSIN3-histone deacetylase complex to the pocket of retinoblastoma tumor suppressor family proteins found in limited discrete regions of the nucleus at growth arrest. *Mol. Cell. Biol.*, **21**, 2918–2932.

20. Wang,X., Nagl,N.G.Jr., Wilsker,D., Van Scoy,M., Pacchione,S., Dallas,P.B. and Moran,E. (2004) Two related ARID family proteins are alternative subunits of human SWI/SNF complexes. *Biochem. J.*, **383**, 319–325.

21. Emery,P., Durand,B., Mach,B. and Reith,W. (1996) RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom. *Nucleic Acids Res.*, **24**, 803–807.

22. Mohrmann,L., Langenberg,K., Krijgsveld,J., Kal,A.J., Heck,A.J.R. and Verrijzer,C.P. (2004) Differential targeting of two distinct SWI/SNF-related Drosophila chromatin-remodeling complexes. *Mol. Cell. Biol.*, **24**, 3077–3088.

23. Lai,A., Marcellus,R.C., Corbeil,H.B. and Branton,P.E. (1999) RBP1 induces growth arrest by repression of E2F-dependent transcription. *Oncogene*, **18**, 2091–2100.

24. Lai,A., Lee,J.M., Yang,W.M., DeCaprio,J.A., Kaelin,W.G.Jr., Seto,E. and Branton,P.E. (1999) RBP1 recruits both histone deacetylase-dependent and -independent repression activities to retinoblastoma family proteins. *Mol. Cell. Biol.*, **19**, 6632–6641.

25. Fleischer,T.C., Yun,U.J. and Ayer,D.E. (2003) Identification and characterization of three new components of the mSin3A corepressor complex. *Mol. Cell. Biol.*, **23**, 3456–3467.

26. Chan,S.W. and Hong,W. (2001) Retinoblastoma-binding protein 2 (Rbp2) potentiates nuclear hormone receptor-mediated transcription. *J. Biol. Chem.*, **276**, 28402–28412.

27. Tan,K., Shaw,A.L., Madsen,B., Jensen,K., Taylor-Papadimitriou,J. and Freemont,P.S. (2003) Human PLU-1 has transcriptional repression properties and interacts with the developmental transcription factors BF-1 and PAX9. *J. Biol. Chem.*, **278**, 20507–20513.

28. Agulnik,A.I., Mitchell,M.J., Lerner,J.L., Woods,D.R. and Bishop,C.E. (1994) A mouse Y chromosome gene encoded by a region essential for spermatogenesis and expression of male-specific minor histocompatibility antigens. *Hum. Mol. Genet.*, **3**, 873–878.

29. Agulnik,A.I., Mitchell,M.J., Mattei,M.G., Borsani,G., Avner,P.A., Lerner,J.L. and Bishop,C.E. (1994) A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human. *Hum. Mol. Genet.*, **3**, 879–884.

30. Balciunas,D. and Ronne,H. (2000) Evidence of domain swapping within the jumonji family of transcription factors. *Trends Biochem. Sci.*, **25**, 274–276.

31. Clissold,P.M. and Ponting,C.P. (2001) JmJC: cupin metalloenzyme-like domains in jumonji, hairless and phospholipase A2beta. *Trends Biochem. Sci.*, **26**, 7–9.

32. Takeuchi,T., Yamazaki,Y., Katoh-Fukui,Y., Tsuchiya,R., Kondo,S., Motoyama,J. and Higashinakagawa,T. (1995) Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. *Genes Dev.*, **9**, 1211–1222.

33. Lee,Y., Song,A.J., Baker,R., Micales,B., Conway,S.J. and Lyons,G.E. (2000) Jumonji, a nuclear protein that is necessary for normal heart development. *Circ. Res.*, **86**, 932–938.

34. Kim,T.G., Kraus,J.C., Chen,J. and Lee,Y. (2003) JUMONJI, a critical factor for cardiac development, functions as a transcriptional repressor. *J. Biol. Chem.*, **278**, 42247–42255.

35. Saikumar,P., Ramachadran,M. and Reddy,P. (1990) Role of the tryptophan repeats and flanking amino acids in Myd-DNA interactions. *Proc. Natl Acad. Sci. USA*, **87**, 8452–8456.

36. Subramaniam,V., Jovin,T.M. and Rivera-Pomar,R.V. (2001) Aromatic amino acids are critical for stability of the bicoid homeodomain. *J. Biol. Chem.*, **276**, 21506–21511.

37. Shaham,S. and Bargmann,C.I. (2002) Control of neuronal subtype identity by the *C. elegans* ARID protein CFI-1. *Genes Dev.*, **16**, 972–983.

38. Kaplan,M.H., Zong,R.T., Herrscher,R.F., Scheuermann,R.H. and Tucker,P.W. (2001) Transcriptional activation by a matrix associating region-binding protein. contextual requirements for the function of bright. *J. Biol. Chem.*, **276**, 21325–21330.

39. Lahoud,M.H., Ristevski,S., Venter,D.J., Jermiin,L.S., Bertoncello,I., Zavarsek,S., Hasthorpe,S., Drago,J., de Kretser,D., Hertzog,P.J. and Kola,I. (2001) Gene targeting of Desrt, a novel ARID class DNA-binding protein, causes growth retardation and abnormal development of reproductive organs. *Genome Res.*, **11**, 1327–1334.

40. Whitson,R.H., Tsark,W., Huang,T.H. and Itakura,K. (2003) Neonatal mortality and leanness in mice lacking the ARID transcription factor Mrf-2. *Biochem. Biophys. Res. Commun.*, **312**, 997–1004.

41. Inoue,H., Furukawa,T., Giannakopoulos,S., Zhou,S., King,D.S. and Tanese,N. (2002) Largest subunits of the human SWI/SNF chromatin-remodeling complex promote transcriptional activation by steroid hormone receptors. *J. Biol. Chem.*, **277**, 41674–41685.

42. Nie,Z., Yan,Z., Chen,E.H., Sechi,S., Ling,C., Zhou,S., Xue,Y., Yang,D., Murray,D., Kanakubo,E., Cleary,M.L. and Wang,W. (2003) Novel SWI/SNF chromatin-remodeling complexes contain a mixed-lineage leukemia chromosomal translocation partner. *Mol. Cell. Biol.*, **23**, 2942–2952.

43. Wang X., Nagl,N.G.Jr., Flowers S., Zweitzig D., Dallas P.B. and Moran E. (2004) Expression of p270 (ARID1A), a component of human SWI/SNF complexes, in human tumors. *Int. J. Cancer*, **112**, 636–642.

44. Shandala,T., Kortschak,R.D. and Saint,R. (2002) The Drosophila retained/dead ringer gene and ARID gene family function during development. *Int. J. Dev. Biol.*, **46**, 423–430.

45. Nixon,J.C., Rajaiya,J. and Webb,C.F. (2004) Mutations in the DNA-binding domain of the transcription factor, Bright, act as dominant negative proteins and interfere with immunoglobulin transactivation. *J. Biol. Chem.*, **279**, 52465–52472.

46. Prochasson,P., Neely,K.E., Hassan,A.H., Li,B. and Workman,J.L. (2003) Targeting activity is required for SWI/SNF function *in vivo* and is accomplished through two partially redundant activator-interaction domains. *Mol. Cell*, **12**, 983–990.

47. Murphy,F.V.,IV and Churchill,M.E. (2000) Nonsequence-specific DNA recognition: a structural perspective. *Structure Fold Des.*, **8**, R83–R89.

48. Thomas,J.O., Travers,A.A. (2001) HMG1 and 2, and related 'architectural' DNA-binding proteins. *Trends Biochem Sci.*, **26**, 167–174.

49. Hsiao,N.W., Samuel,D., Liu,Y.N., Chen,L.C., Yang,T.Y., Jayaraman,G. and Lyu,P.C. (2003) Mutagenesis study on the zebra fish SOX9 high-mobility group: comparison of sequence and non-sequence specific HMG domains. *Biochemistry*, **42**, 11183–11193.

50. Depiereux,E., Baudoux,G., Briffeuil,P., Reginster,I., DeBolle,X., Vinals,C. and Feytmans,E. (1997) Match-Box_server: a multiple sequence alignment tool placing emphasis on reliability. *Comput. Appl. Biosci.*, **13**, 249–256.

51. Bateman,A., Coin,L., Durbin,R., Finn,R.D., Hollich,V., Griffiths-Jones,S., Khanna,A., Marshall,M., Moxon,S., Sonnhammer,E.L.L., Studholme D.J., Yeats C. and Eddy S.R. (2004) The Pfam Protein Families Database. *Nucleic Acids Res.*, **32**, D138–D141.

52. Numata,S., Claudio,P.P., Dean,C., Giordano,A. and Croce,C.M. (1999) Bdp, a new member of a family of DNA-binding proteins, associates with the retinoblastoma gene product. *Cancer Res.*, **59**, 3741–3747.

53. Takeuchi,T., Nicole,S., Misaki,A., Furihata,M., Iwata,J., Sonobe,H. and Ohtsuki,Y. (2001) Expression of SMARCF1, a truncated form of SWI1, in neuroblastoma. *Am. J. Pathol.*, **158**, 663–672.

54. Kozmik,Z., Machon,O., Kralova,J., Kreslova,J., Paces,J. and Vlcek,C. (2001) Characterization of mammalian orthologues of the drosophila osa gene: cDNA cloning, expression, chromosomal-localization, and direct physical interaction with Brahma chromatin-remodeling complex. *Genomics*, **73**, 140–148.

55. Nagase,T., Ishikawa,K., Kikuno,R., Hirosawa,M., Nomura,N. and Ohara,O. (1999) Prediction of the coding sequences of unidentified human genes. XV. The complete sequences of 100 new cDNA

clones from brain which code for large proteins in vitro. *DNA Res.*, **6**, 337–345.

56. Hurlstone,A.F.L., Olave,I.A., Barker,N., van Noort,M. and Clevers,H. (2002) Cloning and characterization of hELD/OSA1, a novel BRG1 interacting protein. *Biochem. J.*, **364**, 255–264.

57. Fattaey,A.R., Helin,K., Dembski,M.S., Dyson,N. and Harlow,E., Vuocolo,G.A., Hanobik,M.G., Haskell,K.M., Oliff,A., Defeo-Jones,D. and Jones,R.E. (1993) Characterization of the retinoblastoma binding proteins RBP1 and RBP2. *Oncogene*, **8**, 3149–3156.

58. Cao,J., Gao,T., Stanbridge,E.J. and Irie,R. (2001) RBP1L1, a retinoblastoma-binding protein-related gene encoding an antigenic epitope abundantly expressed in human carcinomas and normal testis. *J. Natl Cancer Inst.*, **93**, 1159–1165.

59. Lu,P.J., Sundquist,K., Baeckstrom,D., Poulsom,R., Hanby,A., Meier-Ewert,S., Jones,T., Mitchell,M., Pitha-Rowe,P., Freemont,P. and Taylor-Papadimitriou,J. (1999) A novel gene (PLU-1) containing highly conserved putative DNA/chromatin binding motifs is specifically up-regulated in breast cancer. *J. Biol. Chem.*, **274**, 15633–15645.

The DNA-binding properties of the ARID-containing subunits of yeast and mammalian SWI/SNF complexes

Deborah Wilsker, Antonia Patsialou, Steven D. Zumbrun, Suhkmann Kim[1], Yuan Chen[1], Peter B. Dallas[2] and Elizabeth Moran[*]

Fels Institute for Cancer Research, Temple University School of Medicine, Philadelphia, PA, USA [1]Beckman Research Institute of the City of Hope, Duarte, CA, USA and [2]TVW Telethon Institute for Child Health Research and the Center for Child Health Research, Subiaco, Australia

ABSTRACT

SWI/SNF complexes are ATP-dependent chromatin remodeling complexes that are highly conserved from yeast to human. From yeast to human the complexes contain a subunit with an ARID (A-T-rich interaction domain) DNA-binding domain. In yeast this subunit is SWI1 and in human there are two closely related alternative subunits, p270 and ARID1B. We describe here a comparison of the DNA-binding properties of the yeast and human SWI/SNF ARID-containing subunits. We have determined that SWI1 is an unusual member of the ARID family in both its ARID sequence and in the fact that its DNA-binding affinity is weaker than that of other ARID family members, including its human counterparts, p270 and ARID1B. Sequence analysis and substitution mutagenesis reveals that the weak DNA-binding affinity of the SWI1 ARID is an intrinsic feature of its sequence, arising from specific variations in the major groove interaction site. In addition, this work confirms the finding that p270 binds DNA without regard to sequence specificity, excluding the possibility that the intrinsic role of the ARID is to recruit SWI/SNF complexes to specific promoter sequences. These results emphasize that care must be taken when comparing yeast and higher eukaryotic SWI/SNF complexes in terms of DNA-binding mechanisms.

INTRODUCTION

The SWI/SNF complex is an ATP-dependent chromatin remodeling complex that is highly conserved from yeast to human. Most of the subunits of the complex in yeast have recognizable orthologs in higher eukaryotes (reviewed in 1,2). Upon recruitment to specific promoters, the SWI/SNF complex uses the energy of ATP hydrolysis to remodel chromatin. The complex itself binds DNA with high affinity and contains DNA-crosslinking subunits (3,4). From yeast to human, the complex has been found to contain a subunit with an ARID DNA-binding domain. In yeast the subunit is SWI1 and in human there are two alternative subunits: p270 (5,6) and a p270-related protein reported under various names (7–10) and designated ARID1B here in accordance with nomenclature recently approved by both the HUGO Gene Nomenclature Committee (HGNC) (http://www.gene.ucl.ac.uk/nomenclature/) and the Mouse Genomic Nomenclature Committee (MGNC) (http://www.informatics.jax.org/mgihome/nomen/index.shtml). According to this system, the human p270 gene (*SMARCF1*) now has the alternative designation *ARID1A*.

The ARID (A-T rich interaction domain) defines a distinct family of DNA-binding proteins: 15 in human, six in *Drosophila* and two in yeast. They have been found in all eukaryotic organisms studied. ARID family proteins are diverse in function, but all are implicated in the control of cell growth, differentiation or development (reviewed in 11,12). The consensus sequence for the domain extends across 94 amino acid residues and is well conserved. The founding members of the ARID family are *Drosophila* Dri (13) and the closely related mammalian protein Bright (14). Both proteins bind with high affinity to AT-rich sequences, which prompted the naming of the domain. Another mammalian family member, MRF2, shows similar sequence specificity (15). However, sequence-specific DNA binding has not been reported for most ARID proteins. There is clearly variation in ARID family DNA binding behavior, as p270 and its closest *Drosophila* ortholog Osa show no preference for specific sequences (16,17). While examining the DNA-binding properties of the SWI/SNF complex, we have determined that yeast SWI1 is unusual among members of the ARID family in that it binds DNA with much weaker affinity than other ARID proteins, including its human counterparts p270 and ARID1B. We show here the difference in DNA-binding properties between yeast and human ARID subunits of SWI/SNF complexes. The weak affinity of SWI1 for DNA is largely attributable to a gap in the ARID consensus sequence and to the presence of acidic rather than basic residues in the vicinity of a major DNA contact site. SWI/SNF complexes are well

*To whom correspondence should be addressed. Tel: +1 215 707 7313; Fax: +1 215 707 6989; Email: betty@temple.edu

conserved between yeast and humans, so the different DNA-binding behaviors of the respective ARID-containing subunits is unexpected and underscores the greater degree of complexity in the mammalian versions of the complexes.

MATERIALS AND METHODS

Plasmids

GST fusion constructs. The p270 fusion protein is the product of plasmid pNDX (described in 16). The Dri fusion protein is the product of p410 (13), which was kindly provided by R. Saint. The MRF2 fusion protein is the product of pMRF2-GST, which was constructed by ligating a BamHI–SalI restriction fragment from the insert of MRF2pQE30 into the pGEX4T vector (Pharmacia Biotech). The MRF2pQE30 plasmid was described in Yuan *et al.* (18). The SWI1 fusion protein is the product of pSWI1-GST, which was constructed by PCR using plasmid CP623 as template. CP623 was kindly provided by Craig Peterson. Sequence from base pair 1471 to 2399 (numbered according to accession no. X12493) was amplified and cloned into the TOPO vector (Invitrogen) and subcloned into the pGEX4T vector. The translation product extends from residue 264 to 552 according to accession no. P09547 (the SWI1 ARID extends from residue 402 to 492).

In vitro translation constructs. The p270 NE9-B2 *in vitro* expression plasmid contains p270 base pairs 3071–3931 (according to accession no. NM_006015) in the pGEM5Zf(+) vector (Promega). The translation product extends from residue 901 to 1187 (the p270 ARID extends from residue 1013 to 1107). The deletion and substitution mutant plasmids p270ΔL2 and p270ΔL2-DES were constructed in a NE9-B2 background. The p270 pNNE3ΔARID *in vitro* expression plasmid contains p270 base pairs 3071–4505 in the pGEM5Zf(+) vector, with deletion of base pairs 3356–3748. The translation product extends from residue 901 to 1376 with deletion of residues 996–1126. The ARID1B *in vitro* expression plasmid KM15 contains DNA base pairs 2003–3972 generated by RT–PCR from Saos2 cells. The sequence of the entire PCR product was verified according to accession no. NM_020732 and numbered according to the cDNA sequence in accession no. AF253515. The translation product extends from residue 658 to 1313 (the ARID extends from residue 768 to 864). The dead ringer *in vitro* expression plasmid pDriT2 contains Dri sequences expressing residues 258–410 inserted into the pSK-BBV expression vector (the ARID consensus extends from residue 277 to 369). The pSK-BBV vector (described in 19) is a derivative of Bluescript SKII+ engineered to contain black beetle virus ribosome binding sequences to promote more efficient translation *in vitro*. The SWI1 *in vitro* expression plasmid pSWI1.SZ is the TOPO vector construct containing the base pair 1471–2399 PCR fragment described above.

Other plasmids. The pBSII-δ99 plasmid was described previously (20) and was kindly provided by A. Bank.

Generation of p270 amino acid substitution mutations

All mutations were generated using the QuikChange (Stratagene) system according to the manufacturer's instructions. The forward primer used to generate the amino acid substitutions was DES (CCAACCTCAATGTGAGTG-ACGCCAGCTCCTTGGAGAGCCAGTATATCCAG) (substituted bases underlined).

Deletion mutants were generated by a loop-out technique using a primer designed to form a junction between residues at the borders of the deletion. The sequences of the forward primers used to generate the deletions were ΔARID (CCCAAGACAGAATCCAAATCCCAGCCCAAGATCCA-GCCTCC) and ΔL2 (CAACCAACCTCAATGTGAGTGC-TGCCAGCTCCTTG) (nucleotides that mark the boundaries of the loop are underlined).

The sequence changes and the integrity of the surrounding sequences for all mutants were verified by DNA sequencing.

Sequence-specific selection of DNA

GST fusion proteins were used in pull-down assays with the pools of DNA restriction fragments described in the text. The assay was performed as described in Collins *et al.* (17). Restriction fragments were filled in with [α-^{32}P]dATP. Labeled DNA (0.8 μg) was incubated with 100 ng of GST fusion protein bound to glutathione–agarose beads for 1 h at 4°C in λ DNA binding buffer [20 mM HEPES pH 7.6, 1 mM EDTA pH 8, 10 mM (NH$_4$)$_2$SO$_4$, 0.2% Tween-20, 1 mM dithiothreitol (DTT), 25 μg/ml bovine serum albumin (BSA) and 25 μg/ml poly(dI·dC)] plus varying amounts of KCl, as indicated in the text. The beads were washed three times with λ DNA binding buffer minus DTT, BSA and poly(dI·dC). Bound DNA was eluted by boiling in formamide loading buffer [90% formamide, 1× TBE (89 mM Tris base, 89 mM Boric acid, 2 mM EDTA), 0.04% bromophenol blue and 0.04% xylene cyanol], separated on a 6% sequencing gel and visualized by autoradiography.

In vitro translation and DNA cellulose chromatography

The wild-type and mutant plasmid constructs were used to generate [^{35}S]methionine-labeled polypeptides using the TNT coupled reticulocyte system (Promega). *In vitro* translated proteins were diluted in 1 bed volume (0.5 ml) of column loading buffer (10 mM potassium phosphate pH 6.2, 0.5% NP40, 10% glycerol, 1 mM DTT, 1 mg/ml aprotinin, 1 mg/ml pepstatin and 1 mg/ml leupeptin) and applied to native DNA–cellulose columns (Pharmacia). The protein sample was passed through the column four times. Unbound material is designated flow-through (FT). The columns were then washed multiple times with 1.0 bed volume column loading buffer containing 50 mM NaCl (these are the 50 mM wash fractions) and eluted stepwise with column loading buffer adjusted to contain increasing concentrations of NaCl from 100 to 800 mM, as indicated in the text. Fractions were analyzed by SDS–PAGE. The signal on the dried gel was quantified using a phosphoimager (Fuji) and associated software.

RESULTS

p270 binds DNA without sequence specificity

p270 was originally identified as a protein sharing antigenic specificity with p300 and CBP (5,21). Analysis of p270-associated proteins revealed that p270 is a component of human SWI/SNF complexes and determination of the cDNA

sequence suggested that p270 is an ortholog of yeast SWI1 (5,16). The presence of p270 in human SWI/SNF complexes was independently confirmed when a SWI/SNF complex-associated factor designated BAF250 was cloned and yielded a cDNA sequence co-linear with p270 (6). The p270/BAF250 cDNA is now designated the product of the *ARID1A* gene by the Nomenclature Committee of the Human Genome Organization.

We have previously shown by DNA affinity assays and PCR-amplified random oligonucleotide selection that p270 binds duplex DNA with high affinity, but without regard to sequence specificity (16). DNA binding without sequence preference is also a property of Osa, the closest *Drosophila* counterpart of p270 (17). The DNA-binding properties of p270 are illustrated here in a different approach, utilizing natural instead of synthetic DNA. GST fusion proteins were used to probe for preferential binding within a large pool of λ DNA restriction fragments (Fig. 1). Control ARID family proteins Dri and MRF2 show selectivity in this assay, as they did in other approaches (13,15). Increasing the stringency of the interaction by adjusting the salt concentration results in increasingly more specific preference for selected fragments (lanes 5–7 and 8–10). In contrast, a p270 fusion binds the fragments with no obvious selectivity (lanes 2–4). Increasing stringency does not reveal a preference for specific fragments, except for eventual selection of longer fragments over shorter ones, probably because there are more binding surfaces on longer pieces of DNA.

Following our report on the non-selectivity of p270 in the PCR-amplified random oligonucleotide selection assay, Nie *et al.* (6) reported that p270/BAF250 binds selectively in an EMSA assay to a specific pyrimidine-rich sequence. A SWI/SNF-like complex called PYR had previously been identified by its ability to bind this 99 bp stretch of pyrimidine-rich DNA (95% pyrimidine on one strand), which lies between the human fetal and adult β-globin genes (22) and is involved in regulation of the switch from fetal to adult expression. Nie *et al.* (6) proposed that p270/BAF250 is the component responsible for recruiting PYR to the δ-globin gene through its ability to bind the δ99 sequence. Simultaneously, though, the transcription factor Ikaros was identified as the PYR component that binds pyrimidine-rich DNA (20). Ikaros is not an ARID protein and has no detectable relationship to p270. While p270 has not been identified in the PYR complex, there is still the question whether p270 binds preferentially to pyrimidine-rich sequences in a manner that was not detected in the oligonucleotide selection assay. We therefore tested the ability of p270 to select the δ99 sequence from a pool of restriction fragments generated from a δ99-containing plasmid. p270 shows no selectivity for the 110 bp restriction fragment that contains the δ99 sequence (Fig. 2A, lane 2). An alternative restriction digest in which the δ99 sequence is released as part of a 332 bp fragment was also probed. Even with the advantage of greater length, the pyrimidine-rich fragment was not pulled down selectively by p270 (Fig. 2B). We conclude that p270 does not prefer pyrimidine-rich DNA or the δ99 sequence specifically, but in fact binds DNA without regard to sequence. The previously reported preference for this sequence may have been a reflection of the EMSA assay in which a limited range of competing DNA sequences was used to challenge selectivity for the δ99
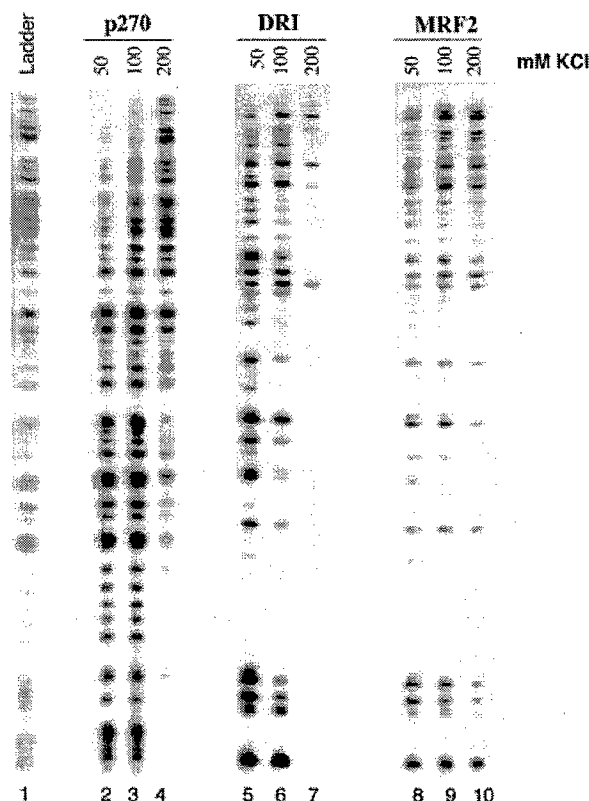

Figure 1. p270 binds DNA non-sequence specifically. λ phage DNA was digested with EcoRI, HindIII and Sau3A1 to generate a large DNA oligonucleotide pool predicted to contain 128 fragments ranging in size from 12 to 2225 bp. The fragments were filled in with [^{32}P]dATP, incubated with GST fusion proteins containing the p270, Dri or MRF2 ARID regions as indicated, pulled down with glutathione beads and analyzed by polyacrylamide gel electrophoresis. Lane 1 shows the unselected pool of DNA fragments. Remaining lanes show the fragments selected in λ DNA binding buffer with increasing KCl concentrations as indicated.

sequence and the competing nucleotide fragments were shorter.

SWI1 has weaker DNA binding affinity than human ARID-containing SWI/SNF subunits

The yeast SWI/SNF complex binds to DNA without sequence specificity (3,23), but the source of the DNA-binding activity in the complex is not well characterized. The SWI1 protein as part of the complex crosslinks to DNA (3,4), but alone has not actually been shown to have DNA-binding activity. When we considered the question of whether SWI1 has sequence specificity we found that SWI1 does not bind well to DNA at all. This is shown in Figure 3. Most of the DNA was released by the 100 mM salt wash and there was no evidence of sequence selectivity. To explore this question and better understand the relationship between yeast and human SWI/SNF complexes, we compared the DNA-binding properties of SWI1 and the ARID-containing subunits of human SWI/SNF complexes. In addition to p270, a partial cDNA product of an independent gene has been identified, originally designated KIAA1235 (24), which has a high degree of overall identity to p270, including the presence of an ARID consensus (7–10).
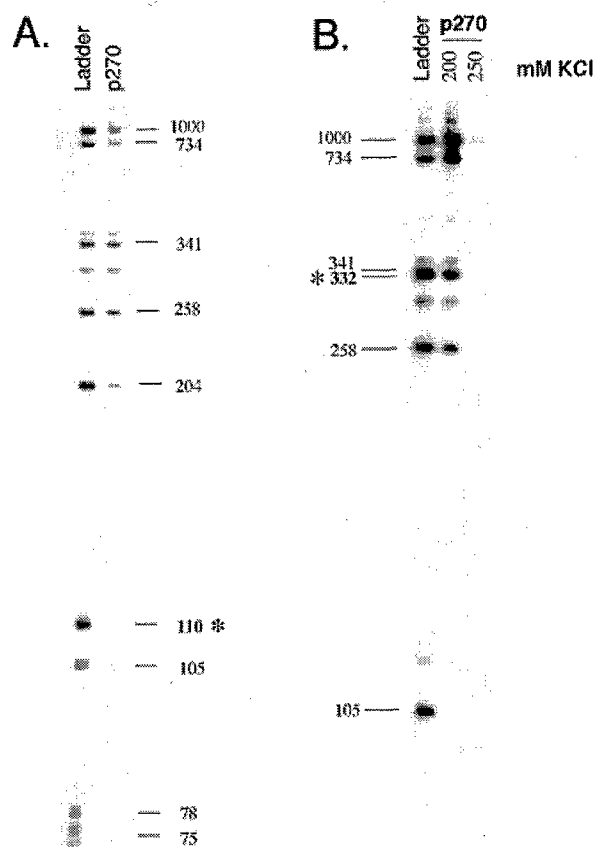
Figure 2. p270 does not bind preferentially to pyrimidine-rich DNA. (A) The pBSII-899 plasmid was digested with EcoR1 and Sau3A1 and labeled with [³²P]dATP to generate a restriction digest ladder as indicated. The 99 bp pyrimidine-rich fragment is contained within a 110 bp fragment indicated by an asterisk. The restriction fragments were incubated with the p270 GST fusion protein in λ DNA binding buffer at 200 mM KCl. (B) The pBSII-899 plasmid was digested with Sau3A1 alone such that the pyrimidine-rich sequence is contained in a 332 bp fragment, indicated by an asterisk. Results from incubations at both 200 and 250 mM KCl are shown.

The Human Genome Organization now recommends that ARID family members carry gene designations that reflect their relationship. According to this scheme, the p270 gene product previously designated *SMARCF1* is designated *ARID1A* and the KIAA1235 gene is designated *ARID1B*. p270 and ARID1B are alternative, mutually exclusive subunits of human SWI/SNF complexes (X. Wang, N.G. Nagl, Jr, M. Van Scoy, S. Pacchione, P.B. Dallas and E. Moran, in preparation). The relationships of p270 and ARID1B to their *Drosophila* and yeast counterparts are shown schematically in Figure 4.

The DNA binding affinity of the yeast and human ARID-containing SWI/SNF components was compared in a DNA-cellulose column chromatography assay, an approach that is unbiased with regard to sequence specificity. ³⁵S-labeled *in vitro* translated proteins were applied to a native DNA-cellulose column and eluted with increasing salt concentrations. The fractions were separated by SDS–PAGE and the protein signal was quantitated by phosphoimager. The signal in each fraction was plotted as a percentage of the total recovered (Fig. 5). What is immediately apparent in this assay
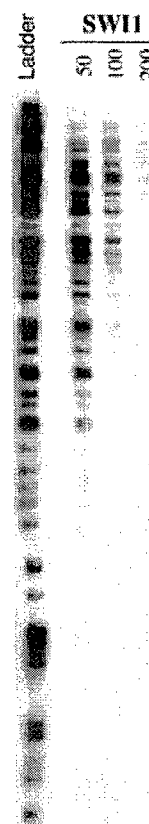
Figure 3. SWI1 binds DNA non-sequence specifically. The DNA-binding activity of a GST fusion protein containing the ARID region of SWI1 was analyzed as described in Figure 1.
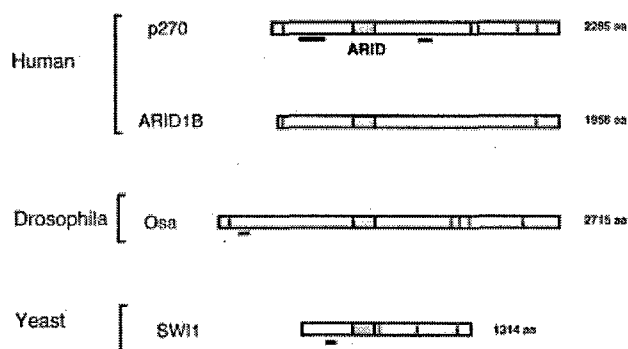

Figure 4. ARID-containing subunits of SWI/SNF complexes in yeast, *Drosophila* and humans. Similar motifs and domains are apparent between the amino acid sequences of yeast SWI1 (accession no. P09547), *Drosophila* Osa (accession no. Q8IN94) and human p270 (accession no. NM_006015) and ARID1B (accession no. AF253515). Yellow boxes denote the ARID, vertical gray lines indicate LXXLL motifs (L symbolizes leucine and X is any amino acid). LXXLL motifs frequently serve as association sites for liganded nuclear hormone receptors (33,34). Horizontal blue bars indicate glutamine-rich (Q-rich) regions. Such regions are implicated in transcriptional activation (see for example 35).

is that p270 and ARID1B show the same high affinity binding as the prototypical ARID protein Dri, but SWI1 has markedly lower affinity for DNA. A control p270 ARID deletion
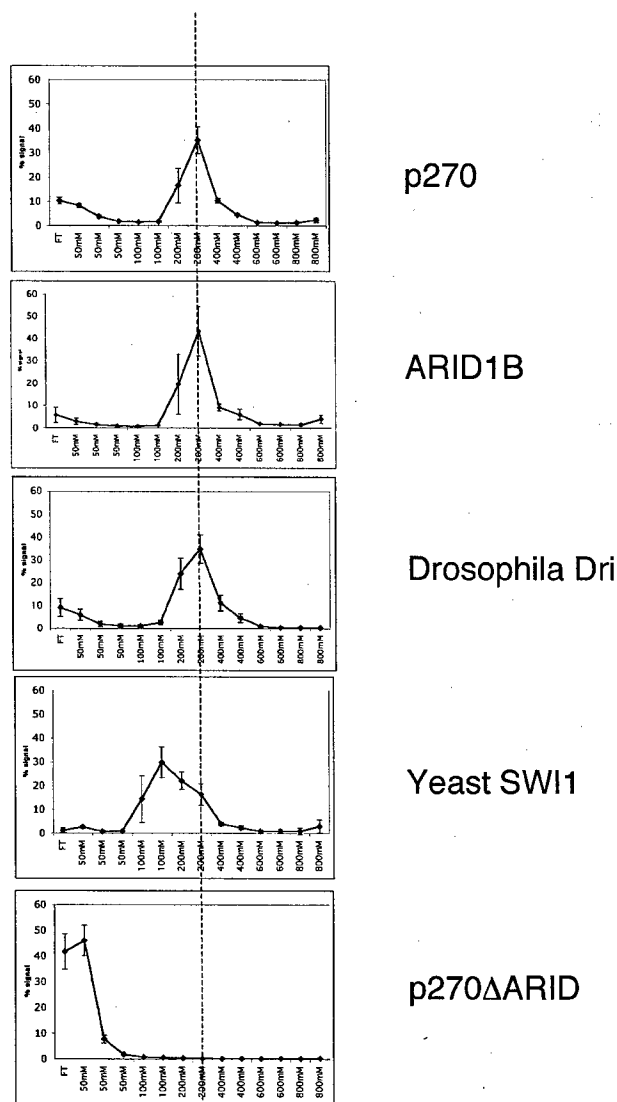
Figure 5. Yeast SWI1 binds DNA poorly compared with other ARID family members, including its human counterparts p270 and ARID1B. *In vitro* translated [^{35}S]methionine-labeled peptides were applied to a native DNA cellulose column as described under Materials and Methods. Bound protein was eluted stepwise with loading buffer adjusted to contain increasing concentrations of NaCl from 100 to 800 mM, as indicated in the figure. Fractions were separated by SDS–PAGE and the p270 signal in each fraction was quantified by phosphoimaging. The results are plotted as the percentage of signal in each fraction relative to the entire signal recovered. Error bars represent the average deviation. Graphs are aligned for ease of comparison. The dashed line indicates the second 200 mM fraction for reference. The proteins analyzed in this experiment were the respective products of plasmids NE9-B2, KM15, p410, pSWI1.SZ and NNE3ΔARID.

construct (p270ΔARID) verifies that the DNA-binding activity observed is a property of the ARID domain.

The SWI1 ARID is poorly conserved in the Loop 2 and H5 region

The ARID is a structurally distinct helix–turn–helix motif based DNA-binding domain. Structures for three ARID proteins have been described: human MRF2, *Drosophila* Dri

and yeast SWI1/ADR6 (18,25–29). The ARID regions of these proteins are aligned in Figure 6. The ARID consensus forms six α-helices (H1–H6). Dri has two additional α-helices (H0 and H7) formed by sequences immediately flanking the consensus. Flexible loops or β-sheets also occur in the structure. NMR studies done on Dri and MRF2 in complex with DNA (25,27) have determined that two regions of the ARID are involved in minor groove and phosphodiester backbone interactions: the Loop 1/β-sheet region and the C-terminus. An interaction point with the major groove was mapped to H5 and the loop preceding it. Sixteen DNA contact residues have been identified in Dri (27); these are indicated by red text in Figure 6. Generally similar contact regions were noted in MRF2, although individual contact residues were not identified (25). NMR of the SWI1 ARID shows that it contains the basic core of six α-helices (28,29). The SWI1 ARID structure was not determined in complex with DNA, so contact residues have not been identified.

A comparison of the SWI1 ARID sequence with p270, Dri and MRF2 does not reveal any obvious deficiency in the predicted minor groove and phosphodiester backbone interaction regions in Loop 1 or the C-terminus of the SWI1 ARID. Basic residues (R and K) are present in positions similar to those seen in p270. However, inspection of the sequence alignments in Figure 6 reveals potentially important differences in the predicted major groove interaction site formed by Loop 2 and H5. A more comprehensive comparison of the Loop 2 and H5 regions is shown in Figure 7. A sequence alignment of SWI1 with all known human, *Drosophila* and yeast ARID family members reveals that SWI1 is a highly unusual member of the family in terms of the length of Loop 2. Loop 2 varies in length by one or two residues among other ARIDs, but is notably shorter in SWI1. SWI1 is also unusual in the distribution of basic and acidic residues in H5. These are indicated by blue and pink shading, respectively in Figure 7. The invariant tyrosine (Y) in H5 is shaded yellow for orientation. A basic residue, R or K, exactly three positions 5' of this tyrosine is nearly invariant and is an identified DNA contact residue in Dri. SWI1 is one of a small subset of ARID proteins that has an acidic residue at or near this position. In the case of SWI1, this is a glutamic acid (E). SWI1 is the only ARID protein known that contains no basic residues between the beginning of Loop 2 and the invariant tyrosine. The lack of positively charged (basic) residues in this region, combined with the presence of negative charges from the acidic residues, very likely contributes to the poor affinity of SWI1 for DNA. The effect of the specific differences between SWI1 and p270 on DNA binding affinity was probed directly as described below.

The sequence differences in Loop 2 and H5 of SWI1 are sufficient to cause defective DNA binding in p270

To evaluate the effect of the specific differences between SWI1 and p270 on DNA binding affinity, site-directed mutagenesis was performed on p270 to mimic the sequence of SWI1. Four residues in Loop 2 of p270, corresponding to the missing residues in SWI1, were deleted. Additionally, three residues in H5 were changed to the corresponding SWI1 residues, as shown in Figure 8. These positions were chosen because they represent the most striking differences in the pattern of basic and acidic residues between the two proteins.
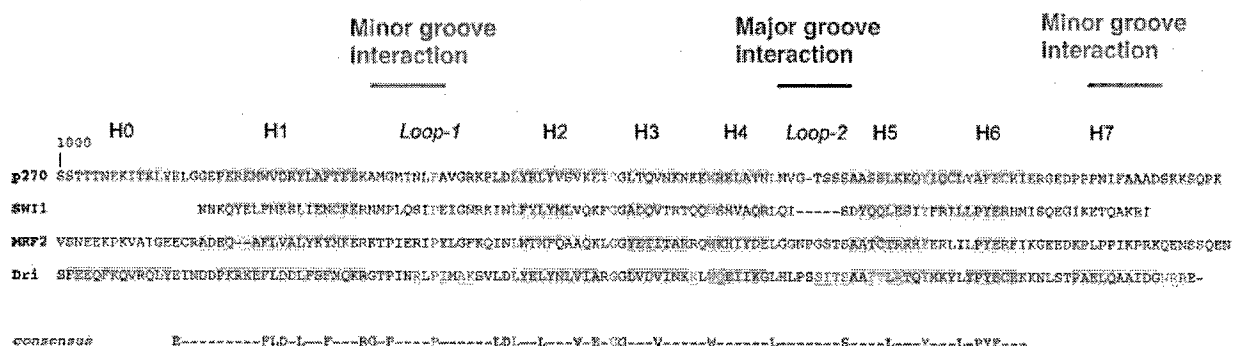
Minor groove interaction   Major groove interaction   Minor groove interaction

HO   H1   Loop-1   H2   H3   H4   Loop-2   H5   H6   H7

p270

SWI1

MRF2

Dri

consensus

Figure 6. Secondary structure of the ARID. The amino acid sequence of the p270 ARID is aligned according to the Clustal W 1.8 multiple sequence alignment program (36) with the corresponding sequences of SWI1, MRF2 and Dri that were used to generate structural data. The computer-generated alignment was modified slightly to reflect higher level structural data. Residue number 1000 (accession no. NM_006015) is indicated in the p270 sequence for reference. The α-helices of each protein are shaded in yellow and numbered above the alignment. The secondary structure of p270 was determined from the backbone resonance assignments obtained recently (37). H5 and H6 in SWI1 are distinguished by a bend between the two adjacent leucines. The ARID consensus forms six α-helices (H1–H6). p270 has an additional short α-helix at the N-terminus and Dri has an extra α-helix on each end (H0 and H7) formed by sequences outside the consensus. Dri also has a β-sheet in place of Loop 1. While the MRF2 and Dri ARID structures differ in significant features, both structures indicate that H5 and Loop 2 contact the major groove and both structures indicate that sequences between H1 and H2 and sequences just downstream of H6 contact the adjacent minor groove and phosphodiester backbone (18,25,26,27). DNA contact residues identified by NMR in Dri (27) are indicated by red text and underlining. The consensus line shows the residues conserved in more than 50% of the 23 ARID family members of human, *Drosophila melanogaster* and *Saccharomyces cereviseae*. Five residues that have proved thus far to be invariant are shown underlined in green.

The p270 mutant constructs were *in vitro* translated and their affinity for DNA was tested by DNA–cellulose chromatography. The wild-type p270 and SWI1 elution profiles are repeated from Figure 5 for ease of comparison. Deletion of four residues from Loop 2 (p270ΔL2) is sufficient to weaken the DNA binding affinity of p270 (Fig. 9). These results are all consistent with the interpretation that Loop 2 makes a significant DNA contact that is lacking in SWI1. It is possible that the length of Loop 2 plays an important role in positioning DNA contact residues in Loop 2 and H5 properly on the DNA. The Loop 2 difference alone, however, is not sufficient to account entirely for the weak DNA binding of SWI1. When the H5 substitutions were introduced into p270 in addition to the Loop 2 deletion (p270ΔL2-DES) the DNA binding affinity of p270 was dramatically impaired and resembled more closely the phenotype of SWI1. The severe effect of these changes demonstrates the significance of the roles of the charged residues in H5. The remainder of the SWI1 ARID sequence may partly compensate for the differences between p270 and SWI1 in the Loop 2/H5 region, as the p270ΔL2-DES mutant is even more severely impaired in this assay than SWI1. Nevertheless, the overall conclusion is that the SWI1 ARID region has a weak DNA binding activity that on its own is not likely to be physiologically significant. Our results indicate that this is an intrinsic feature of the ARID sequence in SWI1, arising from an unusual difference in the length of Loop 2 and from the specific presence of acidic instead of basic amino acid residues at or near a major DNA contact site.

DISCUSSION

SWI1 is an unusual member of the ARID family of DNA-binding proteins. Other ARID family members differ in whether or not their binding is sequence specific, but all family members studied previously show high affinity binding to DNA. The weak DNA binding affinity of the SWI1 ARID is an intrinsic feature of its sequence, arising from specific variations in the major groove interaction site. The human counterparts of SWI1 do, however, bind DNA with an affinity typical of true DNA-binding proteins. This is not the only difference between yeast and human complexes, although the composition and subunit structure of yeast and human complexes are generally well conserved. While the yeast complex has only one ATPase and one ARID-containing protein, human complexes have two alternative ATPase subunits and alternative ARID-containing subunits. Human complexes also have an additional DNA-binding component that has no counterpart in yeast. This is the HMG-containing subunit BAF57, which binds selectively to four-way junction DNA (30). *Drosophila* complexes contain a counterpart to BAF57, designated BAP111 (31), so this is a consistent feature of higher eukaryotes. Human complexes that lack BAF57 are able to bind DNA and remodel chromatin *in vitro* (30), consistent with the fact that there are multiple DNA-binding subunits in the complex.

In spite of the weak DNA-binding activity of SWI1, the yeast SWI/SNF complex does bind DNA with high affinity. In early UV crosslinking experiments, three components were found crosslinked to naked DNA; SWI1 and two other components, p68 and p78, whose DNA-binding properties have not been further characterized (3). In a later study several other members were found to crosslink with nucleosomal DNA (4). These other components may account for the high affinity binding of the complex. The interaction of yeast SWI/SNF complexes with naked DNA is distamycin-sensitive and so appears to occur through minor groove interactions (3,23). This is consistent with our conclusion that the major groove contact region is not functionally conserved in the SWI1 ARID. The complex is not displaced by distamycin when bound to nucleosomes, indicating that other stabilizing interactions occur (23).

Loop 2   Helix 5

```
DRI          LHLPSSIT-SAAFTLRTQYMKYL
MRF2         LGGNPGST-SAATCTRRHYERLI

p270         LNVGTSS--SAASSLKKQYIQCI
KIAA1235     LNVGTSS--SAASSLKKQYIQYL
OSA          LGIGASS--SAAYTLRKHYTKNL
SWI1         LQI-----SDYQOLESIYFRIL

DRIL1        LNLPTSIT-SAAFTLRTQYMKYL
DRIL2        LNLPTSIT-SAAFTLRTQYMKYL
XM_071061    LSLPTTIT-SAAFTLRTQYMKYL
ARID2        FNFPRSCS-NAAFALKQYYLRYL
MRF1         LGGSPGST-SAATCTRRHYERLV
RBP1         LGIPILNS-AASYNLKTAYRKYL
RBP1L1       LGIPVLNS-AAGYNVKCAYKKYL
Jumonji      LRIPKTAQ-ERLAKLQEAYCQYI
SMCX         LNYPPGK--NIGSLLRSHYERIV
SMCY         LHYPPGK--NIGSLLRSHYERII
RBP2         LGYLPGK--GTGSLLKSHYERIL
PLU1         MGFAPGK--AVGSHIRGHYERIL

Yeast YM42   LGYSGRIMSSLSTSLRSAYARIL
Dros CG3654  MCIPKLAQD-RVTKLDDIYCKYL
Dros CG7274  LGFTPCTVS-VMNLVRQAYKKFL
BcDNA GH12174 METLRERCVNGTASIRHIYRRYL
Dros LID     RMQYPSSKS-VGATLKAHYERIL

consensus    L------------LB--Y-B-L
```

Figure 7. Alignment of the Loop 2 and Helix 5 region of human, *Drosophila* and yeast ARID family members. The amino acid sequence extending across the Loop 2 and H5 region of all known human, *Drosophila* and *S.cereviseae* ARID family members are aligned for comparison. *Drosophila* Dri and human MRF2 are shown first to help align their defined H5 and Loop 2. The residues that form H5 are boxed where the structure is known for Dri, MRF2, p270 and SWI1. The ARID-containing members of SWI/SNF complexes (SWI1, *Drosophila* Osa and human p270 and ARID1B) are clustered together. All other mammalian ARID family members are clustered in the third group and the last cluster includes the remaining yeast and *Drosophila* ARID family members. Basic amino acids, arginine (R), lysine (K) and histidine (H), are shaded blue. Acidic amino acids, aspartic acid (D) and glutamic acid (E), are shaded pink. The invariant tyrosine (Y) residue is shaded yellow. Sequences are aligned according to the invariant tyrosine as well as the highly conserved leucine residues that flank the majority of sequences shown. Dashes are inserted where appropriate to maintain the alignment. The consensus line represents residues conserved in at least 50% of the sequences shown. The blue shaded B in the consensus line represents conservation of basic residues at that position.

Yeast SWI1, *Drosophila* Osa and human p270 bind DNA without sequence specificity. Thus, the role of the ARID *per se* is not to recruit the complex to specific promoter elements. The exact biochemical role of the ARID family proteins in SWI/SNF complexes remains to be determined. Deletion of the ARID region of SWI1 does not affect the ability of yeast to grow in conditions that require a functional SWI/SNF complex or the ability of the yeast SWI/SNF complex to remodel nucleosomes *in vitro* (32). Nevertheless, the presence

Loop-2   Helix 5

```
p270         LNVGTSSSAASSLKKQYIQCL
SWI1         LQI----SDYQOLESIYFRIL

p270ΔL2      LNV----SAASSLKKQYIQCL

p270ΔL2-DES  LNV----SDASSLESQYIQCL
```

Figure 8. Mutants generated in p270 to mimic the yeast SWI1 sequence. The Loop 2 and H5 region of wild-type p270 and SWI1 are shown in the top two lines. Four residues, glycine (G), threonine (T) and two serines (S), in Loop 2 of p270 were deleted to create p270ΔL2. To create the mutant p270ΔL2-DES, an alanine (A) and two lysines (K) were changed to the corresponding SWI1 residues, aspartic acid (D), glutamic acid (E) and serine (S). The substituted positions are indicated by black dots. The invariant tyrosine is shaded yellow for reference.
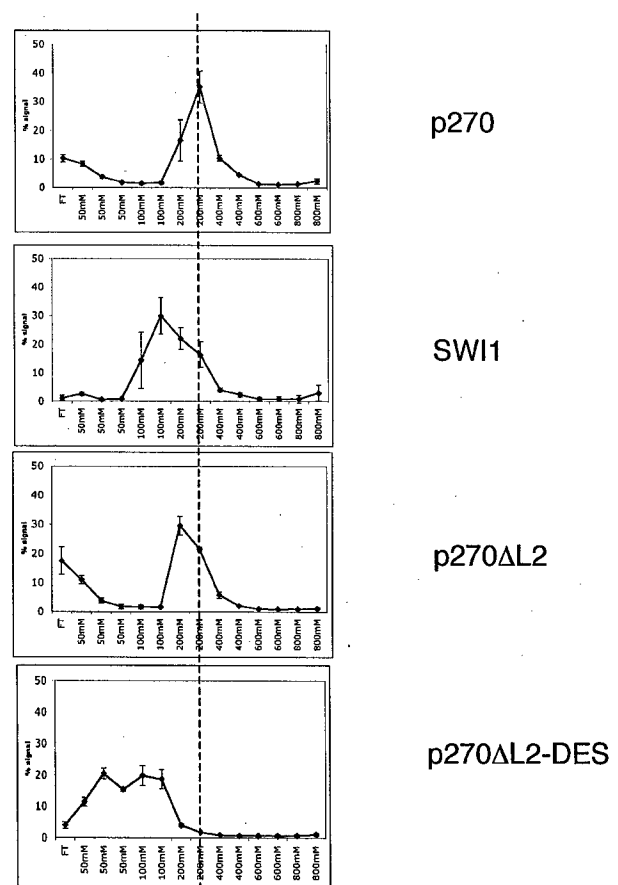

Figure 9. Substitution of SWI1 sequences into p270 is sufficient to make p270 defective for DNA binding. The mutants described in Figure 8 were tested for DNA-binding affinity as described in Figure 4. The wild-type plasmids for p270 and SWI1 (NE9-B2 and pSWI1.SZ) were constructed to generate comparably sized peptides in order to maximize the validity of the comparison. Their elution profiles from Figure 4 are shown again here for ease of comparison. p270ΔL2 and p270ΔL2-DES were constructed in the NE9-B2 background. The dashed line indicates the second 200 mM fraction for reference. Error bars indicate average deviation for at least three experiments. The p270ΔL2-DES elution profile consistently shows two peaks. The reason is not certain, but one possibility is that the accumulated mutations impede proper folding, leading to two populations, one in which structural integrity is severely compromised and another in which the protein has assumed its optimal conformation.

of a functional ARID in the ARID family components of the human and *Drosophila* complexes suggests that this domain has a physiological role in higher eukaryotes. There is experimental evidence that the ARID plays a role in the biological activity of p270 and ARID1B. Deletion of the ARID region of p270 partially reduces its ability to enhance glucocorticoid receptor-mediated transcription in a co-transfection reporter assay (6). Deletion of the ARID from ARID1B abrogates its activity in a similar assay (9). The differences between the complexes in yeast and higher eukaryotes emphasize that caution must be used in making direct comparisons between them.

ACKNOWLEDGEMENTS

REFERENCES

1. Vignali,M., Hassan,A.H., Neely,K.E. and Workman,J.L. (2000) ATP-dependent chromatin-remodeling complexes. *Mol. Cell. Biol.*, **20**, 1899–1910.
2. Martens,J.A. and Winston,F. (2003) Recent advances in understanding chromatin remodeling by Swi/Snf complexes. *Curr. Opin. Genet. Dev.*, **13**, 136–142.
3. Quinn,J., Fyrberg,A.M., Ganster,R.W., Schmidt,M.C. and Peterson,C.L. (1996) DNA-binding properties of the yeast SWI/SNF complex. *Nature*, **379**, 844–847.
4. Sengupta,S.M., Van Kanegan,M., Persinger,J., Logie,C., Cairns,B.R., Peterson,C.L. and Bartholomew,B. (2001) The interactions of yeast SWI/SNF and RSC with the nucleosome before and after chromatin remodeling. *J. Biol. Chem.*, **276**, 12636–12644.
5. Dallas,P.B., Cheney,I.W., Liao,D., Bowrin,V., Byam,W., Pacchione,S., Kobayashi,R., Yaciuk,P. and Moran,E. (1998) p300/CREB binding protein-related protein p270 is a component of mammalian SWI/SNF complexes. *Mol. Cell. Biol.*, **18**, 3596–3603.
6. Nie,Z., Xue,Y., Yang,D., Zhou,S., Deroo,B.J., Archer,T.K. and Wang,W. (2000) A specificity and targeting subunit of a human SWI/SNF family-related chromatin-remodeling complex. *Mol. Cell. Biol.*, **20**, 8879–8888.
7. Kato,H., Tjernberg,A., Zhang,W., Krutchinsky,A.N., An,W., Takeuchi,T., Ohtsuki,Y., Sugano,S., de Bruijn,D.R., Chait,B.T. and Roeder,R.G. (2002) SYT associates with human SNF/SWI complexes and the C-terminal region of its fusion partner SSX1 targets histones. *J. Biol. Chem.*, **277**, 5498–5505.
8. Hurlstone,A.F.L., Olave,I.A., Barker,N., van Noort,M. and Clevers,H. (2002) Cloning and characterization of hELD/OSA1, a novel BRG1 interacting protein. *Biochem. J.*, **364**, 255–264.
9. Inoue,H., Furukawa,T., Giannakopoulos,S., Zhou,S., King,D.S. and Tanese,N. (2002) Largest subunits of the human SWI/SNF chromatin-remodeling complex promote transcriptional activation by steroid hormone receptors. *J. Biol. Chem.*, **277**, 41674–41685.
10. Nie,Z., Yan,Z., Chen,E.H., Sechi,S., Ling,C., Zhou,S., Xue,Y., Yang,D., Murray,D., Kanakubo,E., Cleary,M.L. and Wang,W. (2003) Novel SWI/SNF chromatin-remodeling complexes contain a mixed-lineage leukemia chromosomal translocation partner. *Mol. Cell. Biol.*, **23**, 2942–2952.
11. Wilsker,D., Patsialou,A., Dallas,P.B. and Moran,E. (2002) ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation and development. *Cell Growth Differ.*, **13**, 95–106.
12. Kortschak,R.D., Tucker,P.W. and Saint,R. (2000) ARID proteins come in from the desert. *Trends Biochem. Sci.*, **25**, 294–299.
13. Gregory,S.L., Kortschak,R.D., Kalionis,B. and Saint,R. (1996) Characterization of the dead ringer gene identifies a novel, highly conserved family of sequence-specific DNA binding proteins. *Mol. Cell. Biol.*, **16**, 792–799.
14. Herrscher,R.F., Kaplan,M.H., Lelsz,D.L., Das,C., Scheuermann,C. and Tucker,P.W. (1995) The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. *Genes Dev.*, **9**, 3067–3082.
15. Whitson,R.H., Huang,T. and Itakura.K. (1999) The novel Mrf-2 DNA-binding domain recognizes a five-base core sequence through major and minor-groove contacts. *Biochem. Biophys. Res. Commun.*, **258**, 326–331.
16. Dallas,P.B., Pacchione,S., Wilsker,D., Bowrin,V., Kobayashi,R. and Moran,E. (2000) The human SWI/SNF complex protein, p270, is an ARID family member with nonsequence-specific DNA binding activity. *Mol. Cell. Biol.*, **20**, 3137–3146.
17. Collins,R.T., Furukawa,T., Tanese,N. and Treisman,J.E. (1999) Osa associates with the Brahma chromatin remodeling complex and promotes the activation of some target genes. *EMBO J.*, **18**, 7029–7040.
18. Yuan,Y.C., Whitson,R.H., Liu,Q., Itakura,K. and Chen,Y. (1998) A novel DNA-binding motif shares structural homology to DNA replication and repair nucleases and polymerases. *Nature Struct. Biol.*, **5**, 959–964.
19. Boyd,M.T., Vlatkovic,N. and Haines,D.S. (2000) A novel cellular protein (MTBP) binds to MDM2 and induces a G1 arrest that is suppressed by MDM2. *J. Biol. Chem.*, **275**, 31883–31890.
20. O'Neill,D.W., Schoetz,S.S., Lopez,R.A., Castle,M., Rabinowitz,L., Shor,E., Krawchuk,D., Goll,M.G., Renz,M., Seelig,H.P. *et al.* (2000) An ikaros-containing chromatin-remodeling complex in adult-type erythroid cells. *Mol. Cell. Biol.*, **20**, 7572–7582.
21. Dallas,P.B., Yaciuk,P. and Moran,E. (1997) Characterization of monoclonal antibodies raised against p300: both p300 and CBP are present in intracellular TBP complexes. *J. Virol.*, **71**, 1726–1731.
22. O'Neill,D., Yang,J., Erdjument-Bromage,H., Bornschlegel,K., Tempst,P. and Bank,A. (1999) Tissue-specific and developmental stage-specific DNA binding by a mammalian SWI/SNF complex associated with human fetal-to-adult globin gene switching. *Proc. Natl Acad. Sci. USA*, **96**, 349–354.
23. Cote,J., Peterson,C.L. and Workman,J.L. (1998) Perturbation of nucleosome core structure by the SWI/SNF complex persists after its detachment, enhancing subsequent transcription factor binding. *Proc. Natl Acad. Sci. USA*, **95**, 4947–4952.
24. Nagase,T., Ishikawa,K., Kikuno,R., Hirosawa,M., Nomura,N. and Ohara,O. (1999) Prediction of the coding sequences of unidentified human genes. XV. The complete sequences of 100 new cDNA clones from brain which code for large proteins *in vitro. DNA Res.*, **6**, 337–345.
25. Zhu,L., Hu,J., Lin,D., Whitson,R., Itakura,K. and Chen,Y. (2001) Dynamics of the Mrf-2 DNA binding domain free and in complex with DNA. *Biochemistry*, **40**, 9142–9150.
26. Iwahara,J. and Clubb,R.T. (1999) Solution structure of the DNA binding domain from Dead ringer, a sequence-specific AT-rich interaction domain (ARID). *EMBO J.*, **18**, 6084–6094.
27. Iwahara,J., Iwahara,M., Daughdrill,G.W., Ford,J. and Clubb,R.T. (2002) The structure of the Dead ringer–DNA complex reveals how AT-rich interaction domains (ARIDs) recognize DNA. *EMBO J.*, **21**, 1197–1209.
28. Tu,X., Xiao,Y., Zeng,W. and Shi,Y. (2000) Expression and purification of a recombinant DNA-binding domain of ADR6 protein from *Escherichia coli* and its secondary structure characterization. *Biochim. Biophys. Acta*, **1481**, 167–174.
29. Tu,X., Wu,J., Xu,Y. and Shi,Y. (2001) 1H, 13C and 15N resonance assignments and secondary structure of ADR6 DNA-binding domain. *J. Biomol. NMR*, **21**, 187–188.
30. Wang,W., Chi,T., Xue,Y., Zhou,S., Kuo,A. and Crabtree,G.R. (1998) Architectural DNA binding by a high-mobility-group/kinesin-like subunit in mammalian SWI/SNF-related complexes. *Proc. Natl Acad. Sci. USA*, **95**, 492–498.
31. Papoulas,O., Daubresse,G., Armstrong,J.A., Jin,J., Scott,M.P. and Tamkun,J.W. (2001) The HMG-domain protein BAP111 is important for

the function of the BRM chromatin-remodeling complex *in vivo*. *Proc. Natl Acad. Sci. USA*, **98**, 5728–5733

32. Prochasson,P., Neely,K.E., Hassan,A.H. Li,B. and Workman,J.L. (2003) Targeting activity is required for SWI/SNF function *in vivo* and is accomplished through two partially redundant activator-interaction domains. *Mol. Cell*, **12**, 983–990.

33. Le Douarin,B., Nielsen,A.L., Garnier,J.M., Ichinose,H., Jeanmougin,F., Losson,R. and Chambon P. (1996) A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors. *EMBO J.*, **15**, 6701–6715.

34. Heery,D.M., Kalkhoven,E., Hoare,S. and Parker,M.G. (1997) A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature*, **387**, 733–736.

35. Seipel,K., Georgiev,O. and Schaffner,W. (1992) Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. *EMBO J.*, **11**, 4961–4968.

36. Depiereux,E., Baudoux,G., Briffeuil,P., Reginster,I., DeBolle,X., Vinals,C. and Feytmans,E. (1997) Match-Box_server: a multiple sequence alignment tool placing emphasis on reliability. *Comput. Appl. Biosci.*, **13**, 249–256.

37. Kim,S., Zhu,L., Yu,B. and Chen,Y. (2003) Letter to the editor: Assignment of ^{1}H, ^{13}C and ^{15}N resonances of the ARID domain of p270. *J. Biomol. NMR*, **27**, 227–278.

Review

ARID Proteins: A Diverse Family of DNA Binding Proteins Implicated in the Control of Cell Growth, Differentiation, and Development[1]

Deborah Wilsker, Antonia Patsialou, Peter B. Dallas, and Elizabeth Moran[2]

Fels Institute for Cancer Research and Molecular Biology, Temple University School of Medicine, Philadelphia, Pennsylvania 19140 [D. W., A. P., E. M.], and TVW Telethon Institute for Child Health Research and the Center for Child Health Research, University of Western Australia, Subiaco, Western Australia 6008 [P. B. D.]

Abstract

The ARID family of DNA binding proteins was first recognized ~5 years ago. The founding members, murine Bright and *Drosophila* dead ringer (Dri), were independently cloned on the basis of their ability to bind to AT-rich DNA sequences, although neither cDNA encoded a recognizable DNA binding domain. Mapping of the respective binding activities revealed a shared but previously unrecognized DNA binding domain, the consensus sequence of which extends across ~100 amino acids. This novel DNA binding domain was designated AT-rich interactive domain (ARID), based on the behavior of Bright and Dri. The consensus sequence occurs in 13 distinct human proteins and in proteins from all sequenced eukaryotic organisms. The majority of ARID-containing proteins were not cloned in the context of DNA binding activity, however, and their features as DNA binding proteins are only beginning to be investigated. The ARID region itself shows more diversity in structure and function than the highly conserved consensus sequence suggests. The basic structure appears to be a series of six α-helices separated by β-strands, loops, or turns, but the structured region may extend to an additional helix at either or both ends of the basic six. It has also become apparent that the DNA binding activity of ARID-containing proteins is not necessarily sequence specific. What is consistent is the evidence that family members play vital roles in the regulation of development and/or tissue-specific gene

expression. Inappropriate expression of ARID proteins is also increasingly implicated in human tumorigenesis. This review summarizes current knowledge about the structure and function of ARID family members, with a particular focus on the human proteins.

Introduction

About 5 years ago, a new class of DNA binding proteins, defined by a novel DNA binding domain, was recognized. The two proteins in which this domain was originally defined are murine Bright and *Drosophila* dead ringer. Bright is a B cell-specific transactivator cloned in a search for proteins binding to immunoglobulin heavy-chain matrix-associating regions. Matrix attachment regions are AT-rich sequences, and the Bright protein was indeed found to bind preferentially to AT-rich DNA sequences in an oligonucleotide selection and enhancement protocol (1). At the same time the *Drosophila* gene product, dead ringer (*dri*), was cloned in a search for novel proteins associating with homeobox domains. Homeobox domains are also AT-rich sequences, and the Dri protein was likewise found to bind preferentially to AT-rich DNA sequences in a similar oligonucleotide selection and enhancement protocol (2).

What distinguished both of these proteins at the time was the lack of a recognizable DNA binding domain. When these investigators mapped the DNA binding regions in their respective proteins and realized they had identified highly related sequences, the parameters of a previously unrecognized DNA binding domain became apparent. The degree of conservation in the respective domains is remarkable, given that these proteins were cloned from distantly related organisms and that the proteins are not otherwise similar.

This novel DNA binding domain was designated ARID,[3] based on the shared features of Bright and Dri. The derivation of a Bright/Dri consensus sequence led to the recognition of other ARID-containing proteins already cloned or subsequently added to the database. About a dozen distinct human ARID proteins have been recognized, as well as six *Drosophila* members of the family. ARID proteins or open reading frames are also apparent in yeast, *Arabidopsis*, and *Caenorhabditis elegans*. The majority of ARID-containing proteins were not cloned in the context of DNA binding activity, and their features as DNA binding proteins are only beginning to be investigated.

As awareness of the family has grown, a tighter consensus sequence has emerged. This consensus extends across ~100

[3] The abbreviations used are: ARID, AT-rich interactive domain; MRF, modulation recognition factor; pRb, retinoblastoma protein.

Table 1 Functions of human ARID proteins[a]

Yeast	Drosophila	Mouse	Human	Function
SWI1	Osa (eld)	Osa1	p270	p270 is a component of human SWI/SNF complexes (11, 12, 14) and is deficient in some breast and ovarian cancer lines (60).
			(SMARCF1) (B120) (BAF250)	osa associates with the *Drosophila* brahma (SWI/SNF-related) complex (5, 6), modifies E2F (47), and is an antagonist of wingless (46).
				Swi 1 is a component of the yeast SWI/SNF complex (66).
			KIAA1235	An open reading frame very closely related to p270 across its entire sequence, but with specific modifications in all known functional motifs, including the ARID region (18).
	CG7274		RBP1 (RBBP1)	Retinoblastoma binding protein-1 (20); represses E2F-dependent transcription (22–24).
			RBP1L1 (BCAA)	Retinoblastoma-binding protein 1-like 1. Highly expressed in cancers of various tissue origins but restricted in normal tissue (25).
ORF YMR716w (Ecm5p)	Lid (CG9088)		RBP2 (RBBP2)	Retinoblastoma binding protein-2 (20). Drosophila lid is most homologous to RBP2. Lid was cloned in a screen for new trithorax group genes (7).
			SMCY	An evolutionarily conserved protein encoded on the Y chromosome (35).
			SMCX (XE169)	The X-chromosome homologue of SMCY; SMCX escapes X-inactivation (26).
			PLU-1	Up-regulated in breast cancer (27).
	CG3654		jumonji (JMJ)	Developmentally important in the nervous system, liver, spleen, thymus, and heart (28–30).
			MRF-1	Modulator recognition factor-1; represses CMV enhancer activity (43, 67).
		Desrt	MRF2	Modulator recognition factor-2; represses CMV enhancer activity (43, 67).
	Dri		Bdp (DRIL2)	Bright and dead ringer homologous protein (32).
		Bright	DRIL1 (h Bright)	*dri* is a *Drosophila* gene required for early embryonic patterning (2, 8, 9). Bright is a B cell-specific activator most studied in mouse (1, 68). DRIL-1 is a human product 80% identical to murine Bright. DRIL1 binds the pRb-controlled transcription factor EZF1 and rescues Ras-induced senescence (63).
	BCDNA: GH12174			BCDNA:GH12174 does not appear to have a direct counterpart in human cells.

[a] The 13 human ARID proteins are grouped with their closest *Drosophila* and *S. cerevisiae* counterparts. Alternate protein and gene names for some proteins are indicated in parentheses. Where the murine gene product has been published under a different name, that name is indicated in column 3.

residues, of which ~39 are highly conserved with regard to both identity and spacing. As a point of comparison, the homeodomain consensus spans 60 residues, of which ~20 are highly conserved (reviewed in Ref. 3). The Bright/Dri homology extends ~40 residues past the ARID consensus. This "extended ARID" sequence now appears to be characteristic of just one subfamily. Outside of the ARID region, proteins of the ARID family show diversity of sequence, structure, size, and function, although subgroups are readily discernible. The ARID region itself has proved to be more diverse in structure and function than the highly conserved consensus sequence suggests. The ARID protein family has been reviewed recently (4), but new information has since emerged, particularly in the realization that not all ARID proteins show sequence specificity in their DNA binding activity. The latter point is striking, given the high order of structure and degree of conservation of ARID regions. In general, such features are linked with increasing specificity as seen, for example, in homeodomains. ARID pro-

teins are also becoming increasingly implicated in human tumorigenesis. This review summarizes what is currently known about the salient features of members of the ARID protein family. Our main focus is the mammalian proteins, although the *Drosophila* and yeast proteins are considered where relevant for comparison. Much of the information discussed in the text is summarized in Tables 1 and 2 and Figs. 1 and 2.

Number and Diversity of ARID Proteins

The extensive sequence data now available ensure that the predicted protein repertoires of the well-sequenced organisms are largely known. Two ARID-containing proteins have been revealed in budding yeast. The best known is SWI1, a component of the SWI/SNF complex, a multicomponent complex involved in chromatin remodeling and broad aspects of transcription regulation. The other is an open reading frame homologous to the RBP2/Plu-1/SMCX/SMCY subgroup in humans,

Table 2 Size, location and tissue distribution of the human ARID proteins

	M_r Length[a]	Human chromosome	Tissue distribution
p270 (SMARCF1) (B120) (BAF250)	270,000 2285 aa	1p36.1–p35	Broad. Northern Blots show similar levels of expression in the full range of tissues tested: spleen, thymus, prostate, testis, ovary, small intestine, colon, peripheral blood lymphocytes, heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas (12).
KIAA1235	245,000 1711 aa	6q25.1–q25.3	Broad (18).
RBP1 (RBBP1)	200,000 (observed) 143,000 (predicted) 1257 aa	14q22.3	Broad with some specialization. RBP1 is expressed in all tissues examined by Northern blot, although the level of expression among different tissues is not constant (results with specific tissues were not reported; Ref. 20).
RBP1L1 (BCAA)	1311 aa	1q42.1–q43	Restricted. Among normal tissues, RBP1L1 is well expressed only in testis, but high expression was seen in all cancer tissues examined of breast ovary, lung, colon, and pancreatic origin (25).
RBP2 (RBBP2)	195,000 1722 aa	12p11	Broad with some specialization as indicated for RBP1 (20).
SMCY	1538 aa	Yq11	Specific to males, but RT-PCR indicates similar levels of expression in the full range of tissues tested: brain, kidney, liver, lung, muscle, spleen, and heart (35).
SMCX (XE169)	1560 aa	Xp11.22–p11.21	RT-PCR indicates similar levels of expression in the full range of tissues tested: brain, kidney, liver, lung, muscle, spleen, and heart (35).
PLU-1	1544 aa	1q32.1	Restricted. In normal tissues, Plu-1 is well expressed only in testis, but it is consistently up-regulated in breast cancers (27).
jumonji (JMJ)	160,000 1266 aa	6p24–p23.	Specialized. Abundant in brain, heart, skeletal muscle, kidney, and thymus but hard to detect in lung, liver, or spleen (29).
MRF-1	Unknown	2p11.1	Not reported.
MRF2	83,000 743 aa in mouse	10q11.22	The expression profile of MRF2 is not reported. Expression of Desrt (murine MRF2) is broad with some specialization. A Northern blot shows abundant expression in brain, kidney, and lung; moderate expression in heart, small intestine, and muscle; and no detectable signal in liver, spleen, large intestine, or skin (33).
Bdp (DRIL2)	61,000 560 aa	15q24	RNA was detected in a broad range of tissues but was abundant in placenta, testis, and leukocytes (32).
Bright (DRIL1)	75,000 (observed) 593 aa	19p13.3	Restricted. A ribonuclease protection assay shows message accumulation in mature B cells but not in T cells or immature B cells; in mouse tissues, expression was detected in testis but not in brain, kidney, lung, liver, spleen, or thymus (1).

[a] The total number of amino acids (aa) of the major form of each of the 13 human ARID-containing proteins is shown here. Where endogenous full-length protein has been observed, the relative migration rate (M_r) reported is indicated.

discussed below. Six ARID-containing proteins are apparent in the *Drosophila* genome. Osa is structurally related to SWI1, and associates with the brahma complex, which is the *Drosophila* equivalent of the SWI/SNF complex (5, 6). Little imaginal discs (*lid*) was cloned recently in a screen for new trithorax group genes. It is recognizably similar to the second yeast protein and is closely related to human RBP2 (7). Dri acts as a coactivator or corepressor at specific transcription sites (8, 9) and has no apparent orthologue in yeast. Its closest mammalian counterparts are Bright and Bdp. The *Drosophila* BCDNA:GH12174 open reading frame has an ARID sequence close to Dri and Bright, but without the extended ARID sequences. Two other open reading frames containing the ARID consensus are apparent in the *Drosophila* genome. They are designated CG7274 and CG3654, and outside their ARID regions they are related to the human proteins RBP1 and jumonji, respectively.

The number and forms of ARID proteins broaden further in mammalian cells. Eleven distinct human ARID-containing open reading frames were counted in the Celera human genome sequence (10), but 13 are apparent in public databases. The human ARID-containing proteins vary in size from human Bright (DRIL1) and Bdp, which contain just <600 amino acids, to p270, which contains >2000. p270 is the most direct structural and functional orthologue of SWI1 and Osa. p270 shows ~80% identity with Osa in the well-conserved COOH-terminal region and is stably associated with human SWI/SNF (hSWI/SNF) complexes (11, 12). Four distinct human ARID proteins are very similar to the second ARID protein of yeast. These are RBP2, SMCY, SMCX, and Plu-1, discussed individually below. The most direct human orthologues of Dri are human Bright (DRIL-1) and Bdp (DRIL-2). Human Bright and Bdp are similar in size (M_r 75,000 and M_r 61,000, respectively) and almost iden-
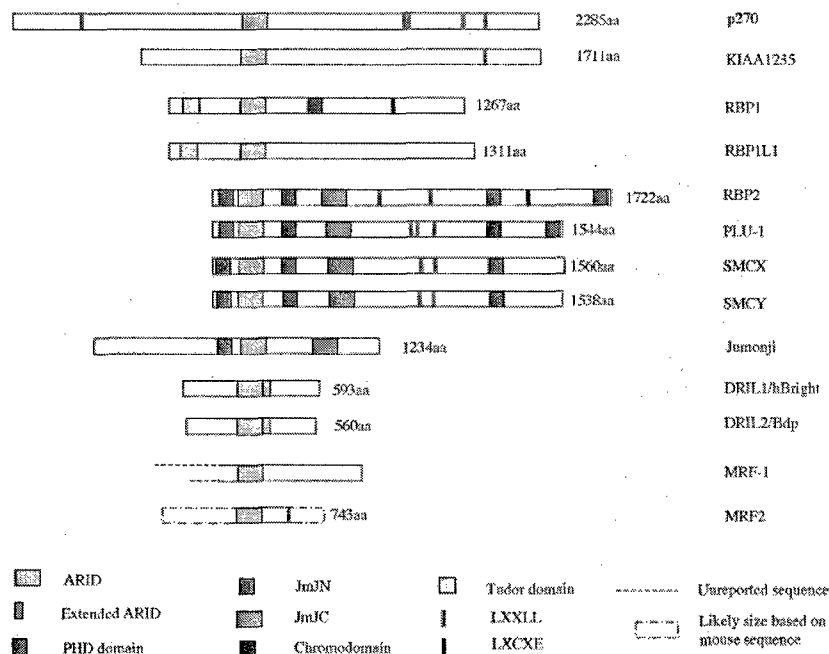
Fig. 1. Schematic representation of the human ARID family proteins. The 13 human ARID family proteins are represented by *open bars* and are aligned according to the position of the ARID sequence (indicated in *yellow*). The relative positions of other well-characterized domains and motifs are represented by differently colored *bars* or *boxes* in the appropriate protein structures and identified at the *bottom* of the figure. The amino acid (*aa*) length of each protein is shown to the *right* of the *bar*. The length of MRF2 is estimated from the corresponding murine product. The full-length sequence of MRF1 is not yet reported. In cases where alternative splice forms are predicted from Genbank sequences, the most complete form is depicted.

Fig. 2. Schematic representation of the *Drosophila* ARID family proteins. The six *Drosophila* ARID family proteins are represented by *open bars* and are aligned according to the position of the ARID sequence (indicated in *yellow*). The relative positions of other well-characterized domains and motifs are represented by differently colored *bars* or *boxes* within the appropriate protein structures and identified at the *bottom* of the figure. The amino acid (*aa*) length of each protein is shown at the *right* of the corresponding *bar*.
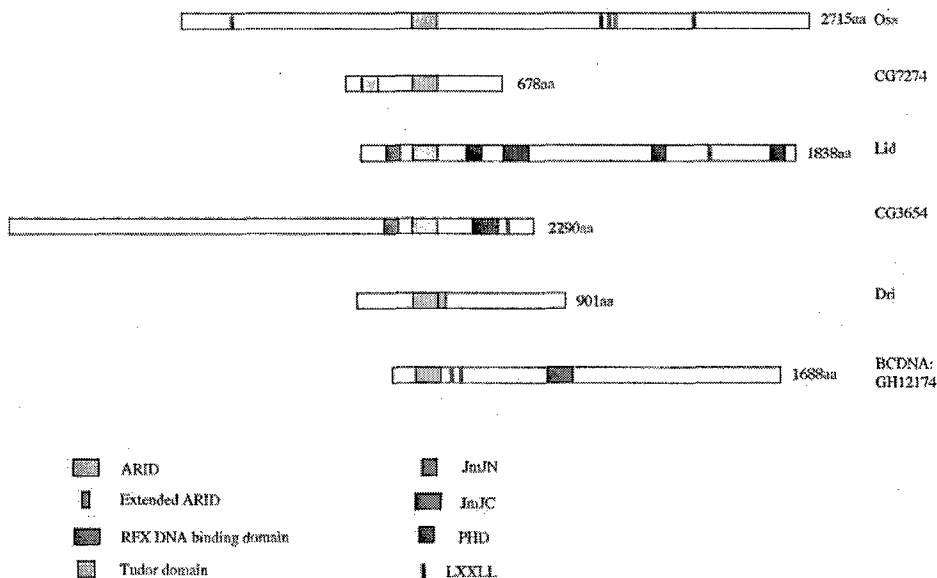
tical in their ARID sequences. Their sequences are not highly similar outside the ARID region, but their relationship to Dri is apparent in the conservation of an extra sequence of ~30 amino acids that extends directly COOH-terminal from the ARID consensus. This extended ARID sequence is >75% identical in Dri, Bright, and Bdp but does not occur in other ARID-containing proteins. Sequences of five other ARID-containing proteins have been identified in the human genome. Each of the human ARID-containing proteins is introduced here briefly and discussed further under specific topics below.

p270. p270 was first recognized and cloned through its shared antigenic specificity with p300 and CBP (11, 13). Immune complex analysis revealed that p270 is an integral member of human SWI/SNF complexes (11, 12). Recently, independent cloning of a band designated BAF250 in hSWI/SNF complexes reaffirmed that BAF250 is indeed p270 (14). p270 was also cloned independently in a screen for expressed sequences containing trinucleotide repeats, although the cDNA sequence reported by these authors contains a frame-shift that results in a predicted molecular weight of only M_r 120,000 (15,

16). The predicted B120 protein has not been unequivocally identified *in vivo*, and RNA probes are assumed here to be detecting p270 expression. A sequence for murine p270 has been reported recently as *Osa*1 (17).

KIAA1235. KIAA1235 was identified in a human fetal brain library in a search for large expressed cDNA sequences (18). The KIAA1235 gene product is very closely related to p270 (>60% identical across its entire sequence), although it is clearly the product of a distinct gene mapping to a different chromosome. Curiously, virtually all known functional motifs, including the ARID sequence, are altered in the KIAA1235 product relative to p270 in ways that suggest the proteins have distinct functions. Analysis with antibodies capable of distinguishing the KIAA1235 protein and p270 indicates that the endogenous KIAA1235 protein migrates at M_r ~245,000 *in vivo* and does associate with hSWI/SNF complexes.[4]

RBP1. RBP1 was cloned in a search for pRb binding partners soon after pRb was identified as a negative regulator of E2F (19, 20). RBP1 contains the LXCXE motif first identified as a pRb binding motif in DNA tumor virus oncogene products such as the adenovirus E1A proteins (reviewed in Ref. 21). RBP1 received relatively little attention until a recent series of reports established that RBP1 acts as a repressor of E2F-dependent transcription and can recruit histone deacetylase activity to pRb/E2F complexes (22–24).

RBP1L1. RBP1-like protein 1 was identified through antibodies to an epitope expressed frequently in human carcinomas. Cloning of the epitope-encoding cDNA revealed a protein that is 40–50% identical to RBP1, although RBP1L1 does not contain an LXCXE motif. RBP1L1 expression is tightly restricted to testis in normal tissues, but expression is abundant in many carcinomas (25).

RBP2. RBP2 was cloned in the same initial screen as RBP1 and also contains an LXCXE (pRb-binding) motif in addition to the ARID consensus (20). The ARID sequence and the LXCXE motif are shared features, but RBP2 is not otherwise related to RBP1. Rather, RBP2 is closely related across its entire length to the SMC proteins and Plu-1. RBP2, Plu-1, and the SMC proteins also share specific sequence motifs with jumonji.

SMCY/SMCX. SMCY was cloned while looking for genes involved in the expression of the minor histocompatibility antigen H-Y. SMCY is encoded on the Y chromosome, and SMCX is the X-chromosome homologue of SMCY. SMCX is one of the few X-chromosome-encoded genes known to escape X inactivation (26). The SMC proteins are closely related to RBP2 but do not contain an LXCXE pRb binding motif.

Plu-1. Plu-1 was identified by differentially screening a fetal brain library with cDNAs prepared from a human mammary epithelial cell line overexpressing c-*ErbB2* in a probe for genes up-regulated in breast cancer. Plu-1 is closely related to RBP2 and the SMC protein, but like the SMC proteins, Plu-1 does not contain the LXCXE pRB binding motif (27). Most human ARID proteins are rather broadly expressed, but Plu-1 expression in normal adult tissue is tightly restricted to testis. In agreement with the method of its isolation, however, Plu-1 is consistently expressed in breast cancers.

jumonji. jumonji was first isolated in a mouse gene trap strategy. In the original study, the mutant *jumonji* gene was linked with formation of an abnormal cruciform-shaped neural groove ("jumonji" translates as "cruciform" in Japanese; Ref. 28). jumonji has since been described as developmentally important in the liver, spleen, thymus, and heart as well as the nervous system (29). The human jumonji sequence is also available (30). In addition to the ARID consensus, jumonji shows significant homology to RBP2 and the SMC proteins in two regions of about 40 and 127 residues. These regions have been respectively designated jmjN and jmjC in a recent report discussing evolutionary relationships among such jumonji domain-containing proteins (31).

Bright (DRIL1). Murine Bright is a B cell-specific transactivator cloned in a search for proteins binding to immunoglobulin heavy-chain matrix-associating regions. Bright and Bdp (see below) are the closest mammalian orthologues of *Drosophila* dead ringer. In addition to the 94-residue ARID consensus common to the entire family, both Bright and Bdp share with dead ringer a highly conserved sequence of ~30 additional residues COOH-terminally extended from the core ARID consensus. The human gene product DRIL-1 (Dri-like protein-1) is 80% identical to murine Bright. Outside of the ARID and the extended ARID sequence, Bright and Bdp are not closely related to each other or to Dri. However, they are distinguished among ARID proteins by their relatively small sizes, with apparent molecular weights in the range of M_r 60,000–75,000.

Bdp (DRIL-2). Bdp was cloned from a human testis library as part of a directed search for a potential tumor suppressor gene (32). Bdp does not appear to be the tumor suppressor sought in the study, but the presence of the ARID consensus prompted further characterization of the protein. Bdp is similar to human Bright in its overall structure but not closely related to Bright outside of the ARID and extended ARID sequences. Bdp and Bright have distinct, only partially overlapping, tissue distribution profiles.

MRF-1. MRF-1 and MRF2 were cloned by virtue of their ability to bind to similar sequences in the transcriptional modulator of the human cytomegalovirus major immediate-early promoter. Although MRF-1 and MRF2 are more closely related to each other in their ARID sequences than they are to other members of the family, they are not related outside of the ARID domain. The entire sequence and molecular weight are not known for either human protein, although the size and sequence of MRF2 can be approximated closely from its mouse counterpart, as discussed below.

MRF2. MRF2 was cloned in the same approach used for MRF-1. The full-length human MRF2 sequence is not yet represented in public databases, but the full-length murine sequence is available as the gene product *desrt* (33). Despite the lack of full-length sequence, MRF2 is one of the best studied of the human ARID proteins in terms of DNA binding activity. A PCR selection and amplification approach shows that the ARID region of MRF2 binds preferentially to an AT-rich core sequence that is similar but not identical to the core sequences recognized by Dri and Bright (34). The three-dimensional structure of the MRF2 ARID region has been

---

[4] X. Wang, N. Nagl. D. Wilsker, and E. Moran, unpublished data.

solved by solution NMR. It differs from that of Dri in several important aspects, discussed further below.

Tissue-specific *versus* Broad Range Expression of Human ARID Proteins

Expression profiles of human ARID proteins range from broad to very narrow. Northern blots indicate that p270 is well expressed in all 16 tissues probed (12). Reverse transcription-PCR results indicate that the p270-related KIAA1235 gene product is also widely expressed in normal tissues (18). Fattaey *et al.* (20) report that RBP1 is expressed in all tissues examined by Northern blot, although the level of expression among different tissues is not constant (results with specific tissues were not reported). Expression of the related protein RBP1L1 in normal human tissues is sharply restricted and abundant only in testis, although it is also abundant in many types of carcinomas. Among the subgroup comprising RBP2, PLU-1, SMCX, and SMCY, three members (RBP2, SMCX, and SMCY) appear to be widely expressed (20, 35), whereas PLU-1 expression is tightly restricted to testis (27). Plu-1 is consistently expressed in breast cancers, however, as discussed further below. Jumonji expression is variable. It is highly expressed in some tissues, such as brain and heart, but not as apparent in others (29). Bright expression is highly restricted. RNase protection assays show an accumulation of Bright message in mature B cells but not in T cells or immature B cells (1). Bdp RNA was detected by Northern blot in a broad range of tissues but was more abundant in placenta, testis, and leukocytes (32). Expression profiles of MRF-1 and MRF2 have not been reported, but Northern blot analysis of Desrt, the murine counterpart of MRF2, shows variable expression across a wide range of tissues (33). Expression patterns for the mammalian ARID proteins are compiled in Table 2.

Structure and DNA Binding Activity of ARID Domains

Not All ARID Proteins Prefer AT-rich Sites. Bright, Bdp, and *Drosophila* Dri are >80% identical in their ARID sequences. The DNA binding behaviors of both Bright and Dri have been well characterized. Oligonucleotide selection and amplification shows that murine Bright has a preference for AT-rich sites similar to those found in the matrix attachment regions that served as the probe for isolation of the protein. The selection technique yielded a core hexamer consensus of (A/G)AT(T/A)AA. The selected sequences also consistently showed ATC runs containing AT dimers, features characteristic of matrix attachment region recognition sites. Nucleotide changes affecting any of these features significantly impacted DNA binding (1). Oligonucleotide selection and amplification analysis of Dri yielded a core hexamer consensus almost identical to that selected by Bright: (A/G)ATTAA (2). This is consistent with the consensus engrailed homeodomain binding site sequence (TCAATTAAATGA) used to isolate Dri. Bdp is able to bind similar matrix attachment sequences as Bright, although the DNA binding activity of Bdp has not yet been explored further (32). Most likely Bright and Bdp perform similar sequence-specific DNA binding functions in different subsets of tissues.

MRF-1 and MRF2 were isolated as proteins binding to AT-rich target sequences in the CMV major immediate-early promoter. The preference of MRF2 for AT rich sites has been demonstrated directly by carbethoxylation interference and in an oligonucleotide selection and amplification assay (34). The consensus preferred binding site is AATA(C/T). The interference assays indicate that MRF2 can distinguish among several similar AT-rich hexamers within the probe, suggesting that other factors in addition to an AT-rich recognition site determine binding specificity. MRF-1 is closely related to MRF2 across the entire ARID region, and unpublished results (cited in Ref. 34) indicate that MRF-1 protects the same sequence as MRF2 in the interference assay.

Exploration of the DNA binding behavior of p270 and Osa has added a new dimension to ARID functions. Oligonucleotide selection and amplification reveals no preference in p270 for AT-rich sequences, and indeed, no identifiable sequence preference at all (12), although in an electromobility shift assay, p270 does appear to interact preferentially with an unusual pyrimidine-rich promoter element in comparison with synthetic oligonucleotides of normal purine/pyrimidine content (14). The general lack of sequence-specific binding in p270 is consistent with a similar finding with Osa in a restriction digest fragment selection assay (6). The behavior of this subset of ARID proteins expands the repertoire of ARID functions, although the physiological role of the non-sequence-specific DNA binding activity in these human and *Drosophila* SWI/SNF complex-associated ARID proteins has not yet been elucidated.

The amino acid sequence of the ARID regions gives no clue to the basis for DNA binding specificity or lack of it. The sequence preferences of Bright and Dri might derive partly from the extended ARID region. However, the activity of MRF2, which does not share the extended ARID sequence, indicates either that the basic ARID consensus is sufficient to specify a preference for AT-rich interactive sites or that non-conserved sequences near the ARID consensus contribute to specificity. The sequence-specific binding activity of Dri and MRF2 is apparent in ARID-containing fragments as small as 152 or 108 amino acid residues, respectively, whereas p270 and Osa fragments as large as 418 or 233 residues, respectively, do not bind DNA specifically.

The DNA binding activity of the remaining ARID family proteins has not been examined in any detail. Thus, we do not really know the full dimensions of the ARID-based DNA binding function. Among the unexplored mammalian proteins, some are highly tissue specific, which may suggest that they, like Bright, are sequence-specific DNA binding proteins. Others, like p270, are widely expressed. A fuller understanding of the family will require a more systematic characterization of individual DNA binding behaviors.

ARID Regions Contact Both Major and Minor Grooves. DNA binding proteins generally recognize their target sequences through base-specific contacts in the major groove (36). For a significant minority, however, sequence recognition occurs primarily through minor groove contacts (37). Because the major and minor groove surfaces of the bases present different chemical substituents, the mechanisms of recognition in each case may be fundamentally different.
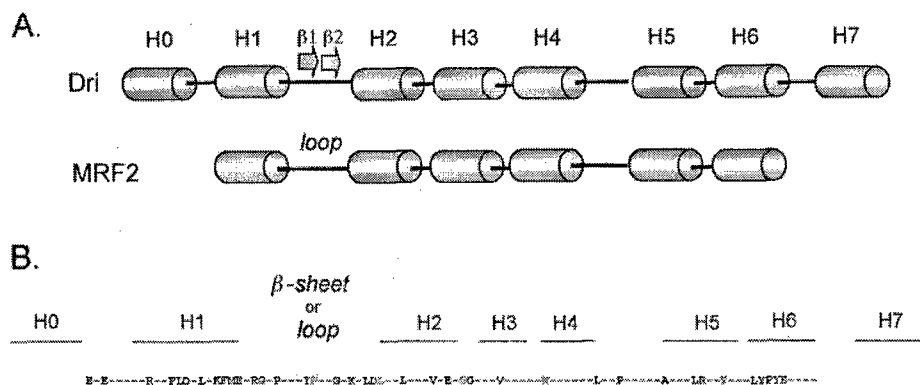
Fig. 3. Alignment of the human ARID consensus sequence with the helical structure of the ARID domain. *A,* the cylinders represent the helical structures of Dri and MRF2. *Arrows,* position of the β-strands in Dri. MRF2 contains a loop at the analogous position. H5 is the predicted recognition helix and is believed to contact the major groove. Other predicted DNA contact regions are the β-sheet and H7 of Dri or the H1-H2 loop and the flexible COOH-terminus of MRF2 (39, 40, 41, 64). *B,* the consensus sequence from the alignment derived in Fig. 4 is shown relative to the ARID helical structure. The sequence is aligned with the Dri helix structure to include the positions of H0 and H7. The five invariant residues in the consensus are underlined and highlighted in red type.

Various lines of evidence suggest that ARID proteins make both major and minor groove contacts.

Several matrix attachment region protein interactions are sensitive to competition by the minor groove-binding antibiotic distamycin A. This sensitivity extends to Bright, indicating that the Bright ARID requires a minor groove interaction to bind its target sequences (1). Distamycin sensitivity suggests that MRF2 also requires a minor groove interaction. However, substitutions in the MRF2-selected core pentamer sequence designed to affect base structure only in the major groove, weaken binding at four of the five positions, suggesting that the nucleotide sequence at most of the core positions is recognized through major groove contacts (34). Required interactions in the minor groove may occur outside the pentamer core.

ARID Structure. The ARID is a highly structured α-helix-based DNA binding domain. Helix-loop-helix and helix-turn-helix motifs each have a two- or three-helix structure in which one helix (the recognition helix) contacts the major groove. Homeodomains have a third helix supporting the alignment of the recognition helix. Other DNA binding protein families that contain helix-turn-helix motifs often have one to three additional conserved helices around the basic motif (reviewed in Ref. 38). A few proteins contact the major groove via β-sheets. Computer algorithms predict that ARID regions consist of a series of at least six α-helices. Nuclear magnetic resonance solution structures have been obtained for two ARID sequences, Dri (39) and MRF2 (40). The structures are similar but differ in important features. MRF2 has six helices (H1 to H6); Dri has these six and one more on each end (H0 and H7) extending beyond the consensus (Fig. 3). MRF2 has a loop between H1 and H2, whereas Dri has a β-sheet located in the analogous position.

Helices H2-H6 form a similar three-dimensional structure in both domains, although they do not superimpose completely. Both structures predict that DNA contact and sequence recognition is made through H5 and its preceeding turn interacting with the major groove of DNA, whereas other residues contact the minor groove or phosphate backbone. Contact with the minor groove is believed to involve the loop between H1 and H2 of MRF2, or the β-sheet in the analogous

position in Dri. The flexible COOH terminus of MRF2 or the equivalent helical structure (H7) in Dri may form additional important contacts with the minor groove or phosphate backbone (41). This is similar in many aspects to the homeodomain interaction with DNA. The recognition helix of the homeodomain makes sequence-specific contact with the major groove, whereas the adjacent minor grooves are contacted, respectively, by a flexible arm and a loop located between the other two helices of the domain (3, 40). However, homeodomains all recognize the same core motif (ATTA; reviewed in Ref. 42), whereas ARIDs do not. In particular, p270 contains a well-conserved ARID consensus, including the predicted recognition helix, but is largely nonspecific in its DNA binding activity (Figs. 3 and 4).

Five residues are absolutely invariable among the human, *Drosophila* and *Saccharomyces cerevisiae* ARID sequences (Fig. 4). The invariable proline (P), trytophan (W), and tyrosine (Y) residues have ring-structured side chains that presumably contribute rigidity to the structure. Basic or polar residues appear at conserved intervals in the consensus sequence and may be instrumental in DNA contact. A few mutagenesis studies have been initiated on the ARID sequence. Deletion of a seven-amino acid stretch that includes the invariable trytophan in Dri abrogates DNA binding and acts in a dominant-negative manner to impair the ability of wild-type Dri to rescue the lethal Dri-null phenotype.[5] A truncation that eliminates H0 and part of H1 impairs the ability of Bright to bind DNA in an electromobility shift assay (1). A truncation that eliminates sequences NH2-terminal to the predicted H1-H2 loop in p270 seriously impairs binding of the p270 ARID to native DNA cellulose columns, as does a combined substitution of the invariable tryptophan and tyrosine residues (12).

Regulation of ARID DNA Binding Activity. There are indications that the DNA binding activity of ARID proteins may be regulated by other cellular processes. MRF2 was isolated as a repressor of the hCMV enhancer, which is repressed in undifferentiated Tera 2 and THP-1 cells. Reti-
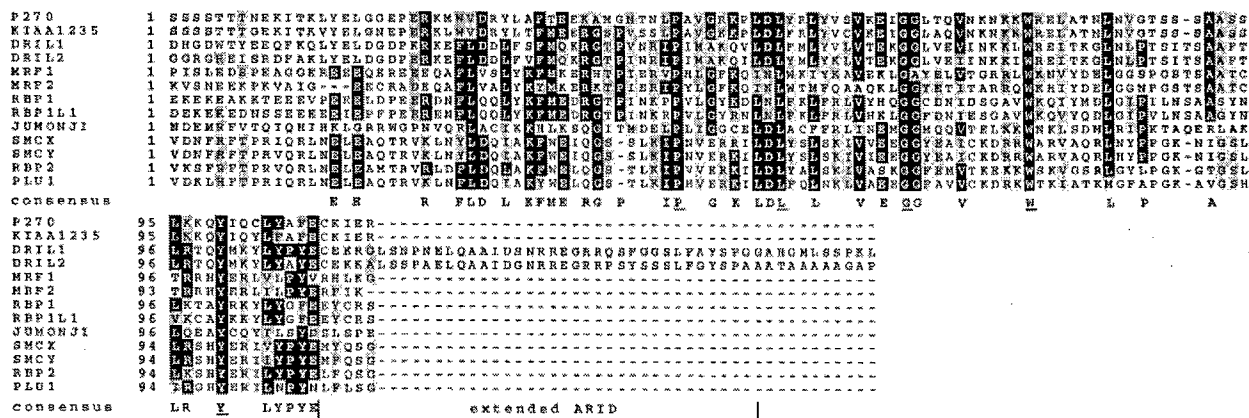
Fig. 4. ARID sequence alignments. The amino acid sequences of the ARID regions of the 13 human ARID-containing proteins were aligned using the Clustal W 1.8 multiple sequence alignment program (65). The *black boxes* indicate residues identical in at least 7 of the 13 proteins. Gray shading indicates positions where at least seven residues are closely related but not identical. Five residues are invariable; these are indicated by *underlining*. The consensus sequence extends across 94 residues, of which 39 are highly conserved with regard to both identity and spacing. (The consensus is defined here as identity at a specific position in at least 7 of the 13 human proteins). The Bright (DRIL1)/Bdp (DRIL2) homology (which is also a feature of *Drosophila* Dri) extends for 35–40 residues past the ARID consensus and appears to be characteristic of one subfamily within the ARID family. Accession numbers for the human gene products used in the alignment are given in parentheses: p270 (AF265208), KIAA1235 (BAA86549), DRIL1 (NP_005215), DRIL2 (NP_006456), MRF-1 (M62324), MRF2 (M73837), RBP1 (P29374), RBP1L1 (NP 057458), Jumonji (Q92833), SMCX (L25270), SMCY (NP_004644), RBP2 (S66431), and PLU-1 (CAB43532).

noic acid-induced differentiation in these cells results in reduced MRF2 DNA binding activity and activation of the enhancer, although the mechanism(s) modifying MRF2 behavior have not been identified, and a direct link between these events has not been established (43). Bright binds DNA as a tetramer, and DNA binding activity is severely impaired if oligomerization is blocked (1). Recently, Bruton's tyrosine kinase was found to be required for Bright DNA binding activity, suggesting links between Bright activity and cell-signaling cascades (Ref. 44; reviewed in Ref. 45).

ARID Protein Functions

Development. Studies of several ARID family members have revealed their importance during development and gene expression. Homozygous null mutants for three of the *Drosophila* ARID family members have been generated, and all are lethal at early stages. Dri-deficient embryos are defective in hindgut and muscle development and in embryonic patterning (9). Dri has been identified specifically as a component of a complex required for dorsal-mediated repression (8). Dri binds AT-rich sites in the 5′ region of the *Drosophila zen* gene, which is called VRR (ventral repression region). Through this specific binding, Dri directs dorsal to this site, and this complex then recruits the repressor, groucho, resulting in ventral repression of *zen*. Of the other *Drosophila* ARID proteins, Osa is required for embryonic segmentation and affects patterning of the wing and imaginal eye disc as well as neuronal differentiation (46). Osa antagonizes signaling by wingless, a Wnt family member, during development (6, 46). Osa is a component of the brahma chromatin remodeling complex (5) and is linked genetically with E2F-mediated transcriptional regulation (47). *lid* is a member of the trithorax group of genes and therefore is predicted to help maintain the expression pattern of homeotic genes at the chromatin level during development (7).

jumonji was the first mammalian ARID gene to be examined in a knockout mouse. *jumonji* homozygous knockouts are embryonic lethal by day E15.5, and the mutant embryos show severe neural tube defects. LacZ expression in *Jmj* transgenic heterozygotes is strong at specific locations in the brain during embryonic development. Postnatally, expression is seen in Purkinje cells and eventually all granule cells. Outside the neural tube, expression is relatively weak throughout development (28).

Desrt homozygous knockout mice show reduced viability. Approximately 50% of homozygotes die *in utero* or within a few hours after birth. Survivors are growth retarded at birth and after, attaining ~69% the weight of their wild-type littermates. *Desrt* heterozygotes have abnormalities in the male and female reproductive organs as well as in the adrenal gland (33). Whole mount *in situ* hybridization shows expression of *Desrt* in the limb bud and interdigital tissue, suggesting a potential role during limb patterning. *Desrt* is also expressed transiently during development in several distinct locales, such as the otic vesicles, endolymphatic diverticulum, auditory meatus, premigratory neural crest, liver diticulum, lung buds, and lining of the oral cavity, suggesting a role in organogenesis (48).

Expression of Osa1, the mouse orthologue of p270, has been examined during mouse development by *in situ* hybridization. Osa1 is ubiquitous in early development but in time becomes more restricted to the limb buds, eye lens, neural tube, and brain (17). However, adult tissue Northern blots indicate that p270 is expressed similarly in all tissues examined (see Table 2).

Bright is primarily expressed in B lymphocytes in adults. Its expression appears to be regulated during fetal development, where it is expressed in pre-B cells and activated mature B lymphocytes. Bright can be detected in the fetal liver, thymus, and brain by reverse transcription-PCR at day

16 of gestation (49). Expression of other mammalian ARID family members has not been examined during development, but their adult tissue distribution patterns are discussed in "Tissue-specific versus Broad Range Expression of Human ARID Proteins" and Table 2. Overall, ARID proteins appear to be widely expressed during development and in some cases are crucial to survival.

Gene Expression. p270 is a component of human SWI/SNF complexes (11). These are ATP-dependent chromatin remodeling complexes first described in yeast. p270 is likely to be involved in recruiting SWI/SNF to specific promoters through interactions with nuclear hormone receptors via its LXXLL motifs. p270 has been shown to bind the glucocorticoid receptor and to activate transcription from a reporter plasmid with glucocorticoid response elements in a hormone-dependent manner (14).

RBP1 associates with the pocket region of pRb and can repress transcription from E2F-dependent promoters. Two repression domains (R1 and R2) have been mapped within RBP1. R2 is COOH-terminal and can associate with the mSin3-HDAC complex (24). Rbp1 recruits this complex to the pRb pocket and can repress E2F-mediated transcription in an HDAC-dependent manner. R1 maps to a region that includes the ARID domain and represses transcription in an HDAC-independent manner (23).

RBP2, like p270, is part of a subset of ARID proteins that contains LXXLL motifs and are therefore predicted to bind nuclear hormone receptors. RBP2 has been found recently to bind the glucocorticoid, estrogen, vitamin D, and retinoic acid receptors *in vitro* and to bind the estrogen receptor *in vivo*. Overexpression of RBP2 can enhance transcription induced by each of these hormones in reporter assays. Addition of pRb further enhanced estrogen-induced transactivation in this assay (50).

SMCY is located on the Y chromosome. Y chromosome genes are generally needed only for male-restricted functions and are expressed only in testes. SMCY, however, is expressed ubiquitously (35). SMCX is located on the X chromosome. X-inactivation in females is believed to be the mechanism by which gene expression from the X-chromosome is equalized between male and female. However, SMCX is one of the few X-chromosome genes known to escape X-inactivation (26). This suggests that the *SMCX* and *SMCY* genes are functional homologues that are largely interchangeable, and that the dose of transcript is important for the function of these genes (51). Although the biological roles of these proteins is not yet clear, SMCY does encode several H-Y antigen epitopes (51).

PLU-1 is the fourth member of the subset of closely related ARID proteins that includes RBP2, SMCX, and SMCY. Little is known about the biological role of PLU-1, but its expression pattern is intriguing. It is tightly restricted to testis in normal tissue blots but up-regulated with high frequency in breast cancer lines (27), as discussed further in "ARID Proteins and Human Tumorigenesis."

Data are beginning to emerge about the biological role of Jumonji. It may be involved in negative regulation of cell growth. Overexpression of jmj in COS and NIH3T3 cells results in decreased cell proliferation (52). Likewise, megakaryocyte progenitor cells from jmj −/− mice show increased cell proliferation but not increased differentiation (53).

MRF2 is a repressor of the hCMV enhancer. It binds the enhancer, which is repressed in undifferentiated Tera 2 and THP-1 cells. Binding activity is markedly reduced in differentiated cells, and enhancer activity is restored (43). Presumably, MRF2 plays a role in the repression of inappropriate differentiation-specific gene expression, but such an activity has not yet been demonstrated directly.

The role of Bright in the regulation of immunoglobulin heavy chain expression has been reviewed recently (45). Although Bright is known to increase immunoglobulin transcription 3–7-fold in antigen-activated B cells, the mechanism of activation is not clear. Bright binds to MARs in the intronic enhancer region of the immunoglobulin heavy chain gene as a tetramer. Bright is able to form bends in DNA of 80–90 degrees and may be able to facilitate long-range interactions in the enhancer region. Bright associates with specific nuclear matrix proteins and may affect chromatin configuration and nuclear sublocalization.

Other Functional Motifs and Domains in ARID Proteins. ARID proteins can be divided into several subgroups based on the presence of additional structural features in the proteins. For the smaller ARID proteins, the ARID consensus is the dominant feature of the protein. Many of the larger ARID proteins have a more complex array of recognized protein domains (see Fig. 1). Jumonji contains two conserved regions that have been recognized in an overlapping family of proteins. These domains are designated JmjN and JmjC, denoting their relative positions within the jumonji protein. The combination of Jmj domains and the ARID domain occurs in *Drosophila* and more distantly related proteins as well. JmjN and JmjC domains occur together in the four ARID proteins of the RBP2/PLU-1/SMCX/SMCY group, and in a few other human proteins linked with transcription, but without ARID domains (31). JmjC domains may occur alone in a much wider group of proteins (54). The functions of the Jmj domains are not yet known, although these authors predict on the basis of certain structural similarities, that JmjC domains may be enzymatically active domains related to cupin metalloenzyme domains. The RBP2/PLU-1/SMCX/SMCY group is linked further by the presence of multiple PHD-type zinc-finger domains. PHD domains are found in >60 human proteins (10).

The RBP2/PLU-1/SMCX/SMCY group also contains LXXLL motifs, which generally serve as binding sites for liganded nuclear hormone receptors (55, 56). These motifs are present in p270 as well. MRF2 and the KIAA1235 protein each have one such motif. The significance of the single motifs is not clear, because functional LXXLL motifs usually occur as multiples. However, MRF2 binding to the human cytomegalovirus enhancer in Tera 2 cells is dependent on whether the cells are treated with retinoic acid (43), implying that the DNA binding activity of MRF2 may be regulated by its binding to the retinoic acid receptor. Binding to nuclear hormone receptors has been demonstrated directly in p270, which binds the glucocorticoid receptor *in vitro* and *in vivo* (14), and in RBP2, which binds several receptors *in vitro* and the estrogen receptor *in vivo* (50). RBP2 and p270 can both increase hormone-responsive activation of reporter plasmids, as discussed above.

RBP1 and RBP1L1 are related across their entire length and share a Tudor domain near each NH_2 terminus. Nine Tudor domain-containing proteins have been identified in the human genome (10). The Tudor domain is found in many proteins that colocalize with ribonucleoprotein or single-strand DNA-associated complexes in the nucleus, the mitochondrial membrane, or at kinetochores. One of these is the *SMN* gene, defects in which cause spinal muscular atrophy. The Tudor domain mediates binding of the SMN protein to spliceosomal core proteins (57).

Although generally similar, RBP1 and RBP1L1 differ in important aspects. Unique motifs in RBP1 include a partial (69% complete) chromodomain and an LXCXE motif, which specifies pRb binding. An LXCXE motif is, however, present in RBP2, which was isolated as a pRb binding protein. Bdp may also contain a pRb binding activity. Bdp does not have an LXCXE motif, and the pRb binding activity observed was impaired by substitution of the invariable proline or trytophan in the ARID sequence, implying it is dependent on the integrity of the ARID structure (32). This interaction has not been demonstrated *in vivo*.

ARID Proteins and Human Tumorigenesis

Emerging data indicate that aberrant expression of ARID proteins is a fairly common feature of mammalian tumor cells. Plu-1 was cloned directly as a product that is specifically up-regulated in breast tumor cells (27). Plu-1 is well expressed in at least four of five common breast cancer cell lines examined but poorly expressed in at least six of eight colon cancer lines (27). RBP1L1 was also cloned directly as a tumor antigen. Similar to Plu-1, its expression in normal human tissue is abundant only in testis, but RBP1L1 expression is abundant in all types of carcinomas screened: breast, ovary, lung, colon, and pancreatic. A link between high expression in human cancers and in normal testis has been noted before and is discussed in Cao *et al.* (25).

Components of human SWI/SNF complexes are frequently lost or altered in tumor cells (58, 59), and this pattern is now known to include the ARID-containing p270 as well. Reduced expression of p270 has been observed in 3 of 21 common breast cancer lines screened as well as in C33A cervical carcinoma cells (14, 60). Another example of ARID loss in tumorigenesis is the *SMCY* gene, which is lost with high frequency in prostate tumor samples (61).

Two different ARID-containing proteins (RBP1 and RBP2) were cloned through association with pRb. Although their expression has not been screened in tumor cells, they clearly have the potential to contribute to control of cell proliferation. Bdp also shows some evidence of a pRb binding function (32). DRIL1 binds to the E2F transcription factor (62) and has very recently been shown to rescue Ras-induced senescence in primary murine fibroblasts and cause them to become oncogenic (63). No link has yet been made between tumorigenesis and the PLU-1 subfamily protein SMCX or the remaining ARID proteins, jumonji, MRF-1, and MRF2.

Overview

The ARID region appears as a single-copy motif in 13 human proteins, ranging in size from human Bright (DRIL1) and Bdp, which contain just <600 amino acids, to p270, which contains >2000. Most of the human ARID proteins can be viewed as a series of pairs or a group of four, with distinct but clearly related sequences and/or additional shared motifs that distinguish them from other members of the ARID family. The ARID consensus spans ~100 residues, of which 30–40 are highly conserved in proteins from the full range of eukaryotic species. The domain has a more complex structure than most other α-helix-based DNA binding domains. The two ARID structures thus far available predict that major groove contact is made through a specific α-helix in a core structure similar to the helix-turn-helix motif in homeodomains. However, the ARID sequence can form at least six α-helices (as in the case of MRF2) and as many as eight as seen in Dri. Moreover, Dri also contains a two-stranded β-sheet, which in this case appears to contact the minor groove. p270 and *Drosophila* Osa each contain a well-conserved ARID consensus, including the predicted recognition helix, but both appear to be largely nonspecific in their DNA binding activity.

Although the ARID family is smaller than most other families of DNA binding proteins, it nevertheless encompasses both ubiquitously expressed members and members whose expression is highly restricted. Gene regulation activities among the ARID proteins also cover a broad range. For example, Bright plays a particular role at matrix attachment regions, RBP1 is part of a deacetylase-associated repressor complex, and p270 is a component of nucleosome remodeling complexes. The ARID sequence has apparently been adopted by a range of proteins with diverse DNA binding needs, but this sense of broad adaptation contrasts with the highly conserved sequence of the motif and with the conservation of the full range of ARID protein structure and function from fly to human. Continued analysis of ARID proteins should shed new light on the nature of DNA protein interactions.

Acknowledgments

References

1. Herrscher, R. F., Kaplan, M. H., Lelsz, D. L., Das, C., Scheuermann, R., and Tucker, P. W. The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. Genes Dev., *9:* 3067–3082, 1995.

2. Gregory, S. L., Kortschak, R. D., Kalionis, B., and Saint, R. Characterization of the *dead ringer* gene identifies a novel, highly conserved family of sequence-specific DNA binding proteins. Mol. Cell. Biol., *16:* 792–799, 1996.

3. D'Elia, A. V., Tell, G., Paron, I., Pellizzari, L., Lonigro, R., and Damante, G. Missense mutations of human homeoboxes: a review. Hum. Mutat., *18:* 361–374, 2001.

4. Kortschak, R. D., Tucker, P. W., and Saint, R. ARID proteins come in from the desert. Trends Biochem. Sci., *25:* 294–299, 2000.

5. Vázquez, M., Moore, L., and Kennison, J. A. The trithorax group gene *osa* encodes an ARID-domain protein that genetically interacts with the brahma chromatin-remodeling factor to regulate transcription. Development (Camb.), *126:* 733–742, 1999.

6. Collins, R. T., Furukawa, T., Tanese, N., and Treisman, J. E. *osa* associates with the Brahma chromatin remodeling complex and promotes the activation of some target genes. EMBO J., *18:* 7029–7040, 1999.

7. Gildea, J. J., Lopez, R., and Shearn, A. A screen for new trithorax group genes identified little imaginal discs, the Drosophila melanogaster homologue of human retinoblastoma binding protein 2. Genetics, 156: 645–663, 2000.

8. Valentine, S. A., Chen, G., Shandala, T., Fernandez, J., Mische, S., Saint, R., and Courey, A. J. Dorsal-mediated repression requires the formation of a multiprotein repression complex at the ventral silencer. Mol. Cell. Biol., 18: 6584–6594, 1998.

9. Shandala, T., Kortshak, R. D., Gregory, S., and Saint, R. The Drosophila dead ringer gene is required for early embryonic patterning through regulation of argos and buttonhead expression. Development (Camb.), 126: 4341–4349, 1999.

10. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. The sequence of the human genome. Science (Wash. DC), 291: 1304–1351, 2001.

11. Dallas, P. B., Cheney, I. W., Liao, D., Bowrin, V., Byam, W., Pacchione, S., Kobayashi, R., Yaciuk, P., and Moran, E. p300/CREB binding protein-related protein p270 is a component of mammalian SWI/SNF complexes. Mol. Cell. Biol., 18: 3596–3603, 1998.

12. Dallas, P. B., Pacchione, S., Wilsker, D., Bowrin, V., Kobayashi, R., and Moran, E. The human SWI/SNF complex protein, p270, is an ARID family member with nonsequence-specific DNA binding activity. Mol. Cell. Biol., 20: 3137–3146, 2000.

13. Dallas, P. B., Yaciuk, P., and Moran, E. Characterization of monoclonal antibodies raised against p300: both p300 and CBP are present in intracellular TBP complexes. J. Virol., 71: 1726–1731, 1997.

14. Nie, Z., Xue, Y., Yang, D., Zhou, S., Deroo, B. J., Archer, T. K., and Wang, W. A specificity and targeting subunit of a human SWI/SNF family-related chromatin-remodeling complex. Mol. Cell. Biol., 20: 8879–8888, 2000.

15. Takeuchi, T., Bing-Kun, C., Qiu, Y., Sonobe, H., and Ohtsuki, Y. Molecular cloning and expression of a novel human cDNA containing CAG repeats. Gene (Amst.), 204: 71–77, 1997.

16. Takeuchi, T., Furihata, M., Heng, H. H. Q., Sonobe, H., and Ohtsuki, Y. Chromosomal mapping and expression of the human B120 gene. Gene (Amst.), 213: 189–193, 1998.

17. Kozmik, Z., Machon, O., Kralova, J., Kreslova, J., Paces, J., and Vlcek, C. Characterization of mammalian orthologues of the Drosophila osa gene: cDNA cloning, expression, chromosomal localization, and direct physical interaction with brahma chromatin-remodeling complex. Genomics, 73: 140–148, 2001.

18. Nagase, T., Ishikawa, K., Kikuno, R., Hirosawa, M., Nomura, N., and Ohara, O. Prediction of the coding sequences of unidentified human genes. XV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. DNA Res., 6: 337–345, 1999.

19. Defeo-Jones, D., Huang, P. S., Jones, R. E., Haskell, K. M., Vuocolo, G. A., Hanobik, M. G., Huber, H. E., and Oliff, A. Cloning of cDNAs for cellular proteins that bind to the retinoblastoma gene product. Nature (Lond.), 352: 251–254, 1991.

20. Fattaey, A. R., Helin, K., Dembski, M. S., Dyson, N., Harlow, E., Vuocolo, G. A., Hanobik, M. G., Haskell, K. M., Oliff, A., Defeo-Jones, D., and Jones, R. E. Characterization of the retinoblastoma binding proteins RBP1 and RBP2. Oncogene, 8: 3149–3156, 1993.

21. Beck, G. R., Jr., Zerler, B., and Moran, E. Introduction to DNA tumor viruses: adenovirus, simian virus 40, and polyomavirus. In: D. J. McCance (ed.), Human Tumor Viruses, pp. 51–86. Washington, D.C. ASM Press, 1998.

22. Lai, A., Marcellus, R. C., Corbeil, H. B., and Branton, P. E. RBP1 induces growth arrest by repression of E2F-dependent transcription. Oncogene, 18: 2091–2100, 1999.

23. Lai, A., Lee, J. M., Yang, W-M., DeCaprio, J. A., Kaelin, W. G., Jr., Seto, E., and Branton, P. E. RBP1 recruits both histone deacetylase-dependent and -independent repression activities to retinoblastoma family proteins. Mol. Cell. Biol., 19: 6632–6641, 1999.

24. Lai, A., Kennedy, B. K., Barbie, D. A., Bertos, N. R., Yang, X. J., Theberge, M. C., Tsai, S. C., Seto, E., Zhang, Y., Kuzmichev, A., Lane, W. S., Reinberg, D., Harlow, E., and Branton, P. E. RBP1 recruits the mSIN3-histone deacetylase complex to the pocket of retinoblastoma tumor suppressor family proteins found in limited discrete regions of the nucleus at growth arrest. Mol. Cell Biol. 21: 2918–2932, 2001.

25. Cao, J., Gao, T., Stanbridge, E. J., and Irie, R. RBP1L1, a retinoblastoma-binding protein-related gene encoding an antigenic epitope abundantly expressed in human carcinomas and normal testis. J. Natl. Cancer Inst., 93: 1159–1165, 2001.

26. Agulnik, A. I., Mitchell, M. J., Mattei, M. G., Borsani, G., Avner, P. A., Lerner, J. L., and Bishop, C. E. A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human. Hum. Mol. Genet., 3: 879–884, 1994.

27. Lu, P. J., Sundquist, K., Baeckstrom, D., Poulsom, R., Hanby, A., Meier-Ewert, S., Jones, T., Mitchell, M., Pitha-Rowe, P., Freemont, P., and Taylor-Papadimitriou, J. A novel gene (PLU-1) containing highly conserved putative DNA/chromatin binding motifs is specifically upregulated in breast cancer. J. Biol. Chem., 274: 15633–15645, 1999.

28. Takeuchi, T., Yamazaki, Y., Katoh-Fukui, Y., Tsuchiya, R., Kondo, S., Motoyama, J., and Higashinakagawa, T. Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. Genes Dev., 9: 1211–1222, 1995.

29. Lee, Y., Song, A. J., Baker, R., Micales, B., Conway, S. J., and Lyons, G. E. Jumonji, a nuclear protein that is necessary for normal heart development. Circ. Res., 86: 932–938, 2000.

30. Berge-Lefranc, J. L., Jay, P., Massacrier, A., Cau, P., Mattei, M. G., Bauer, S., Marsollier, C., Berta, P., and Fontes, M. Characterization of the human *jumonji* gene. Hum. Mol. Genet., *5:* 1637–1641, 1996.

31. Balciunas, D., and Ronne, H. Evidence of domain swapping within the jumonji family of transcription factors. Trends Biochem. Sci., *25:* 274–276, 2000.

32. Numata, S., Claudio, P. P., Dean, C., Giordano, A., and Croce, C. M. Bdp, a new member of a family of DNA-binding proteins, associates with the retinoblastoma gene product. Cancer Res., *59:* 3741–3747, 1999.

33. Lahoud, M. H., Ristevski, S., Venter, D. J., Jermiin, L. S., Bertoncello, I., Zavarsek, S., Hasthorpe, S., Drago, J., de Kretser, D., Hertzog, P. J., and Kola, I. Gene targeting of Desrt, a novel ARID class DNA-binding protein, causes growth retardation and abnormal development of reproductive organs. Genome Res., *11:* 1327–1334, 2001.

34. Whitson, R. H., Huang, T., and Itakura, K. The novel Mrf-2 DNA-binding domain recognizes a five-base core sequence through major and minor-groove contacts. Biochem. Biophys. Res. Commun., *258:* 326–331, 1999.

35. Agulnik, A. I., Mitchell, M. J., Lerner, J. L., Woods, D. R., and Bishop, C. E. A mouse Y chromosome gene encoded by a region essential for spermatogenesis and expression of male-specific minor histocompatibility antigens. Hum. Mol. Genet., *3:* 873–878, 1994.

36. Pabo, C. O., and Sauer, R. T. Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem., *61:* 1053–1095, 1992.

37. Bewley, C. A., Gronenborn, A. M., and Clore, G. M. Minor groove-binding architectural proteins: structure, function, and DNA recognition. Annu. Rev. Biophys. Biomol. Struct., *27:* 105–1031, 1998.

38. Luscombe, N. M., Austin, S. E., Berman, H. M., and Thornton, J. M. An overview of the structures of protein-DNA complexes. Genome Biol., *1:* 1–10, 2000.

39. Iwahara, J., and Clubb, R. T. Solution structure of the DNA binding domain from Dead ringer, a sequence-specific AT-rich interaction domain (ARID). EMBO J., *18:* 6084–6094, 1999.

40. Yuan, Y-C., Whitson, R. H., Liu, Q., Itakura, K., and Chen, Y. A novel DNA-binding motif shares structural homology to DNA replication and repair nucleases and polymerases. Nat. Struct. Biol., *5:* 959–964, 1998.

41. Zhu, L., Hu, J., Lin, D., Whitson, R., Itakura, K., and Chen, Y. Dynamics of the Mrf-2 DNA-binding domain free and in complex with DNA, Biochemistry, *40:* 9142–9150, 2001.

42. Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wuthrich, K. Homeodomain-DNA recognition. Cell, *78:* 211–223, 1994.

43. Huang, T. H., Oka, T., Asai, T., Okada, T., Merrills, B. W., Gertson, P. N., Whitson, R. H., and Itakura, K. Repression via differentiation-specific factor of the human cytomegalovirus enhancer. Nucleic Acids Res., *24:* 1695–1701, 1996.

44. Webb, C. F., Yamashita, Y., Ayers, N., Evetts, S., Paulin, Y., Conley, M. E., and Smith, E. A. The transcription factor Bright associates with Bruton's tyrosine kinase, the defective protein in immunodeficiency disease. J. Immunol., *165:* 6956–6965, 2000.

45. Webb, C. F. The transcription factor, Bright, and immunoglobulin heavy chain expression. Immunol. Res., *24:* 149–161, 2001.

46. Treisman, J. E., Luk, A., Rubin, G. M., and Heberlein, U. eyelid antagonizes wingless signaling during Drosophila development and has homology to the Bright family of DNA binding proteins. Genes Dev., *11:* 1949–1962, 1997.

47. Staehling-Hampton, K., Ciampa, P. J., Brook, A., and Dyson, N. A genetic screen for modifiers of E2F in *Drosophila melanogaster*. Genetics, *153:* 275–287, 1999.

48. Ristevski, S., Tam, P. P., Kola, I., and Hertzog, P. Desrt, an AT-rich interaction domain family transcription factor gene, is an early marker for nephrogenic mesoderm and is expressed dynamically during mouse limb development. Mech. Dev., *104:* 139–142, 2001.

49. Webb, C. F., Smith, E. A., Medina, K. L., Buchanan, K. L., Smithson, G., and Dou, S. Expression of bright at two distinct stages of B lymphocyte development. J. Immunol., *160:* 4747–4754, 1998.

50. Chan, S. W., and Hong, W. Retinoblastoma-binding protein 2 (Rbp2) potentiates nuclear hormone receptor-mediated transcription. J. Biol. Chem., *276:* 28402–28012, 2001.

51. Agulnik, A. I., Longepied, G., Ty, M. T., Bishop, C. E., and Mitchell, M. Mouse H-Y encoding *Smcy* gene and its X chromosomal homolog Smcx. Mamm. Genome, *10:* 926–929, 1999.

52. Toyoda, M., Kojima, M., and Takeuchi, T. Jumonji is a nuclear protein that participates in the negative regulation of cell growth. Biochem. Biophys. Res. Commun., *274:* 332–336, 2000.

53. Kitajima, K., Kojima, M., Kondo, S., and Takeuchi, T. A role of *jumonji* gene in proliferation but not differentiation of megakaryocyte lineage cells. Exp. Hematol., *29:* 507–514, 2001.

54. Clissold, P. M., and Ponting, C. P. JmjC: cupin metalloenzyme-like domains in jumonji, hairless and phospholipase A2β. Trends Biochem. Sci., *26:* 7–9, 2001.

55. Le Douarin, B., Nielsen, A. L., Garnier, J-M., Ichinose, H., Jeanmougin, F., Losson, P., and Chambon, P. A possible involvement if TIFa and TIFb in the epigenetic control of transcription by nuclear receptors. EMBO J., *15:* 6701–6715, 1996.

56. Heery, D., Kalkhoven, M. E., Hoare, S., and Parker, M. G. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. Nature (Lond.), *387:* 733–736, 1997.

57. Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., and Sattler, M. SMN tudor domain structure and its interaction with the Sm proteins. Nat. Struct. Biol., *8:* 27–31, 2001.

58. Klochendler-Yeivin, A., Fiette, L., Barra, J., Muchardt, C., Babinet, C., and Yaniv, M. The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression. EMBO Rep., *1:* 500–506, 2000.

59. Muchardt, C., and Yaniv, M. When the SWI/SNF complex remodels the cell cycle. Oncogene, *20:* 3067–3075, 2001.

60. DeCristofaro, M. F., Betz, B. L., Rorie, C. J., Reisman, D. N., Wang, W., and Weissman, B. E. Characterization of SWI/SNF protein expression in human breast cancer cell lines and other malignancies. J. Cell. Physiol., *186:* 136–145, 2001.

61. Perinchery, G., Sasaki, M., Angan, A., Kumar, V., Carroll, P., and Dahiya, R. Deletion of Y-chromosome specific genes in human prostate cancer. J. Urol., *163:* 1339–1342, 2000.

62. Suziki, M., Okuyama, S., Okamoto, S., Shirasuna, K., Nakajima, T., Hachiya, T., Nojima, H., Sekiya, S., and Oda, K. A novel EZF binding protein with Myc-type HLH motif stimulates E2F-dependent transcription by forming a heterodimer. Oncogene, *17:* 853–865, 1998.

63. Peeper, D. S., Shvarts, A., Brummelkamp, T., Douma, S., Koh, E. Y., Daley, G. Q., and Bernards, R. A functional screen identifies *hDRIL1* as an oncogene that rescues RAS-induced senescence. Nat. Cell Biol., *4:* 148–153, 2002.

64. Iwahara, J., Wojciak, J. M., and Clubb, R. T. Improved NMR spectra of a protein-DNA complex through rational mutagenesis and the application of a sensitively optimized isotope-filtered NOESY experiment. J. Biomol. NMR, *19:* 231–241, 2001.

65. Depiereux, E., Baudoux, G., Briffeuil, P., Reginster, I., De Bolle, X., Vinals, C., and Feytmans, E. Match-Box_server: a multiple sequence alignment tool placing emphasis on reliability. Comput. Appl. Biosci., *13:* 249–256, 1997.

66. Peterson, C. L., and Tamkun, J. W. The SWI-SNF complex: a chromatin remodeling machine? Trends Biochem. Sci., *20:* 143–146, 1995.

67. Thrower, A. R., Bullock, G. C., Bissell, J. E., and Stinski, M. F. Regulation of a human cytomegalovirus immediate-early gene (US3) by a silencer-enhancer combination. J. Virol., *70:* 91–100, 1996.

68. Wang, Z., Goldstein, A., Zong, R. T., Lin, D., Neufeld, E. J., Scheuermann, R. H., and Tucker, P. W. Cux/CDP homeoprotein is a component of NF-muNR and represses the immunoglobulin heavy chain intronic enhancer by antagonizing the bright transcription activator. Mol. Cell. Biol., *19:* 284–295, 1999.

1
2
3
4
5
6
7 **The p270 (*ARID1A/SMARCF1*) subunit of mammalian SWI/SNF-related complexes is**
8 **essential for normal cell cycle arrest.**
9
10

11 Norman G. Nagl, Jr.[1], Antonia Patsialou[1], Dale S. Haines[1], Peter B. Dallas[1†], George R. Beck,
12 Jr.[2], and Elizabeth Moran[1*]
13
14 [1]Fels Institute for Cancer Research and Molecular Biology, Temple University School of Medicine,
15 Philadelphia PA 19140; [2]Laboratory of Cancer Prevention, National Cancer Institute, Frederick,
16 MD 21702
17
18
19
20
21
22

23 <u>Running Title:</u> Role of p270 (*ARID1A*) in cell cycle regulation.

24
25 <u>Key Words:</u> p270, SWI/SNF, BAF, BRG1, hSNF5, p21, cdc2, siRNA, ARID1A, ARID1B,
26 SMARCF1, cell cycle, ARID family
27

28 †current address: Institute for Child Health Research and the Center for Child Health Research,
29 University of Western Australia, Subiaco, Western Australia 6008.
30
31

32

33 Address correspondence to:

34 Elizabeth Moran, Ph.D.,
35 Fels Institute for Cancer Research and Molecular Biology,
36 Temple University School of Medicine, Philadelphia, PA 19140.
37 Ph: (215) 707-7313
38 Fax: (215) 707-6989
39 Email: betty@temple.edu

1 **ABSTRACT**
2
3   Mammalian SWI/SNF-related complexes are ATPase-powered nucleosome remodeling

4 assemblies crucial for proper development and tissue specific gene expression. The ATPase

5 activity of the complexes is critical for tumor suppression in mice and humans. The complexes

6 also contain seven or more noncatalytic subunits, only one of which, hSNF5/Ini1/BAF47, has

7 been individually identified as a tumor suppressor thus far. The noncatalytic subunits include

8 p270/ARID1A, which is of particular interest because recent results from a cDNA tissue array

9 analysis and corroborating screens of panels of tumor cell lines indicate p270 may be deficient in

10 as many as 30% of renal carcinomas and 10% of breast carcinomas. The complexes can also

11 include an alternative ARID1B subunit, which is closely related to p270, but the product of an

12 independent gene. The respective importance of p270 and ARID1B in the control of cell

13 proliferation was explored here using an siRNA approach and a cell system that permits analysis

14 of differentiation-associated cell cycle arrest. The p270-depleted cells fail to undergo normal cell

15 cycle arrest upon induction, as evidenced by continued synthesis of DNA. These lines fail to

16 show other characteristics typical of arrested cells, including up-regulation of p21, down-

17 regulation of cyclins, and decreases in histone H4 expression and cdc2-specific kinase activity.

18 The requirement for p270 is evident separately in both the up-regulation of p21 and the down-

19 regulation of E2F-responsive products. In contrast, the ARID1B-depleted lines behaved like the

20 parental cells in each of these assays. These results show that p270-containing complexes are

21 functionally distinct from ARID1B-containing complexes. They provide a direct biological basis to

22 support the implication from tumor tissue screens that deficiency of p270 plays a causative role in

23 carcinogenesis.

1

2 INTRODUCTION

3

4   The ATPase-powered SWI/SNF chromatin remodeling complex in yeast regulates the

5   mating type switch and other areas of specialized gene expression (reviewed in 1). Mammalian

6   SWI/SNF-related complexes likewise contain an ATPase-powered nucleosome remodeling

7   activity associated with transcriptional regulation. The activity of the complexes is crucial for

8   proper tissue specific gene expression, development, and hormone responsiveness (reviewed in

9   1). More recently it has become apparent that these complexes also play critical roles in

10 suppression of tumorigenesis in mice and humans (reviewed in 2).

11

12   The complexes contain seven or more noncatalytic subunits that presumably help to

13 modulate the targeting and activity of the ATPase. Mammalian complexes have variable

14 compositions because some subunits occur as sets of related proteins. For example, there are

15 two alternative ATPase subunits: mammalian BRM and BRG1. These are closely related

16 proteins, but in mouse knockout studies only BRG1 proved essential for embryonic development

17 and tumor suppression (3; 4). The ATPases are mutated in multiple human tumor cell lines and

18 their loss correlates with poor prognosis of non-small cell lung cancers (5; 6; 7). Noncatalytic

19 components of the complex may be important for tumor suppression as well; however, their

20 individual roles are less well understood.

21

22   Among the noncatalytic subunits, hSNF5 (syns: INI1; BAF47) is recognized as a tumor

23 suppressor in mice (8; 9; 10). In humans, hSNF5 is deficient in a high proportion of pediatric

24 malignant rhabdoid tumors (e. g. 11; 12; 13). Germ-line mutations have been identified, and

25 carriers are pre-disposed to malignant rhabdoid tumors and tumors of the central nervous system

26 (14; 15; 16).

27

1   Expression of functional BRG1 or hSNF5 is associated with specific aspects of cell cycle

2 regulation. Expression of the cell cycle inhibitor p21$^{CIP1/WAF1}$ has been repeatedly identified as

3 BRG1-responsive, and several studies indicate that BRG1-dependent or hSNF5-dependent cell

4 cycle arrest is enacted through a pRb-dependent or overlapping pathway (e. g. 17; 18; 19; 20; 21;

5 22; 23; 24; 25). However, these effects have only been seen in the context of re-introduction of

6 exogenous complex components into tumor cell lines where they were deficient. The significance

7 of the complexes in the expression of these biological targets has yet to be demonstrated during

8 differentiation-associated cell cycle arrest, when the effects of complex dysfunction on

9 carcinogenesis would be most significant.

10

11   Subunits required for the tumor suppression activity of the complexes have great

12 potential as diagnostic and prognostic markers, and as targets for drug therapy. Thus, a major

13 question now is the distinction of which additional noncatalytic subunits are required for the cell

14 cycle arrest functions of the complexes. The noncatalytic components of the complex include the

15 p270 subunit (26; 27; 28) (syns.: ARID1A, SMARCF1, BAF250a, hOSA1), which is a member of

16 the ARID family of DNA binding proteins (reviewed in 29, 30). The role of p270 in cell cycle

17 regulation is of particular interest because recent results from a cDNA tissue array analysis and

18 corroborating screens of panels of tumor cell lines, indicate p270 may be deficient in as many as

19 30% of renal carcinomas and 10% of breast carcinomas (28; 31; 32). A mutually exclusive

20 alternative to p270 in the complexes is the ARID1B (syns: hOSA2, BAF250b) subunit, which is

21 approximately 50% identical to p270 across its entire length, but is the product of an independent

22 gene. The ARID family proteins are determinants that distinguish key divisions among the

23 multiple, distinct SWI/SNF-related complexes that exist in mammalian cells. A major distinction is

24 between the complex first identified as the BRG1-associated factors (BAF) complex (also called

25 the human SWI/SNF or hSWI/SNF complex) and a distinct complex designated PBAF. The BAF

26 complex contains at least BRG1 (or BRM), p270 (or ARID1B), BAF170, BAF155, BAF60, BAF57,

27 BAF53, actin, and hSNF5. The PBAF complex is characterized by the absence of both p270 and

4

1   ARID1B and the presence of a 180 kDa protein designated Polybromo (syn: BAF180). Thus,

2   p270 and ARID1B distinguish between the BAF and PBAF complexes, while BRG1 and hSNF5

3   do not (subunit composition of SWI/SNF-related complexes is reviewed in 1). In addition to the

4   BAF and PBAF division, the BAF series of complexes itself encompasses at least four different

5   entities because p270 and ARID1B can each associate with mammalian BRM and BRG, in all

6   four possible combinations (28; 33; 34), so that p270 and ARID1B each define a specific limited

7   set among the various combinational permutations of SWI/SNF-related complexes that exist in

8   mammalian cells.

9

10   The importance of p270 and ARID1B in proliferation control was explored here using an

11   siRNA approach. The knockdowns were constructed in the MC3T3-E1 pre-osteoblast line

12   because these non-transformed cells undergo a tightly regulated and well-characterized

13   progression through cell cycle arrest and into tissue-specific gene expression (e.g. 26; 35; 36; 37;

14   38; 39). This is an important model system because it permits an examination of the normal roles

15   of the complex subunits during differentiation-associated cell cycle arrest. Parental MC3T3-E1

16   cells arrest by day 4 post-induction with the differentiation signal. The results described here

17   show in parallel conditions that p270-depleted cells fail to arrest normally. This is evidenced by

18   continued synthesis of DNA, and by a lack of other characteristics typical of arrested cells,

19   including up-regulation of p21, down-regulation of cyclins, and decreases in histone H4

20   expression and cdc2-specific kinase activity. The ARID1B-depleted lines behaved like the

21   parental cells in each of these assays. The analysis of the respective roles of p270 (ARID1A) and

22   ARID1B establishes a new paradigm that the choice of ARID-containing subunits confers

23   specificity of function on the complexes. The specific requirement for p270 is evident separately

24   in both the up-regulation of p21 and the down-regulation of E2F-responsive products. The clinical

25   findings suggested indirectly that deficiency of p270 plays a causative role in carcinogenesis.

26   The identification of specific proliferation control steps dependent on the presence of p270 now

27   provides a direct molecular basis to support the clinical findings. Moreover, the demonstration

5

1   that the complexes are required separately for regulation of at least two distinct steps in

2   proliferation control underscores the carcinogenic potential of cells that have lost function of a

3   required subunit.

4

5

MATERIALS AND METHODS

Materials. FBS was purchased from Summit Biotech (Fort Collins, CO), a-MEM from Irvine Scientific (Santa Ana, CA), and penicillin and streptomycin from Mediatech (Herndon, VA). Histone H1, Ascorbic acid, β-glycerol phosphate, and protease inhibitors were obtained from Sigma Chemical Co. (St. Louis, MO), and G418 from Gibco BRL (Grand Island, NY). Radiochemicals were obtained from NEN.

siRNA and isolation of stable p270 knockdown lines. The siRNA sequences were tested in a pSUPER vector constructed as described in (40). Test oligonucleotides were synthesized as complementary pairs, each 64 bases long, containing two inversely repeated copies of a 19 base pair target sequence separated by a 9 base pair spacer region. Six different target sequences were tested in transient expression assays in 293T cells against an exogenously introduced p270 partial expression construct. In a similar manner, four different target sequences were tested against an ARID1B partial expression construct. Expression was monitored by Western Blot. The most effective sequences were chosen for construction of the stable knockdown lines. The pSUPER-derived vectors containing the respective knockdown sequences (pSUPER.p270.7182 and pSUPER.ARID1B.5400) were introduced into MC-3T3-E1 cells by lipofection together with a selectable neo marker. G418-resistant clones were amplified and screened by Western blot for p270 expression. Aliquots of low passage depleted lines were frozen as stocks. The target sequences for each construct were designed against nucleotide stretches that are identical between the mouse and human genes so that they can be used in cells of either species origin.

Cell Culture. Low passage MC3T3-E1 cells were a gift from Roland Baron (Yale University, New Haven, CT). Cells were maintained in α-MEM plus 10% fetal bovine serum supplemented with 50 U per ml penicillin and 50 μg per ml streptomycin. For differentiation assays, cells were plated at an approximate initial density of 5 X 10^4 cells per cm^2. Differentiation was induced by addition

7

1   of 50 μg per ml final concentration ascorbic acid and 10 mM final concentration β-glycerol

2   phosphate to standard growth medium. The medium was changed every 3-4 days and the

3   inducing agents were replaced with each media change.

4

5   **Immunoblotting.**   Cells were washed and harvested in PBS and lysed in p300 lysis buffer

6   [0.1% Nonidet P-40, 250 mM sodium chloride, 20 mM sodium phosphate (pH 7.0), 30 mM sodium

7   pyrophosphate, 5 mM dithiothreitol and protease and phosphatase inhibitors: 0.1 mM sodium

8   vanadate, 1 mM phenylmethylsulfonyl flouride (PMSF), 100 kIU aprotinin, 1 ug/ml leupeptin, and

9   1 ug/ml pepstatin]. Proteins were separated by polyacrylamide gel electrophoresis, transferred to

10   Immobilon-P membrane (Millipore), and visualized as described previously (41).

11

12   **Radiolabeling and immunoprecipitation**. Cells were washed with methionine-free and serum-

13   free α-MEM and incubated with this medium for one hour. 200 μCi of [^{35}S]-methionine (Perkin-

14   Elmer, Boston MA or Amersham, Piscataway NJ) was added to each 10 cm monolayer, and the

15   plates were incubated for a further three hours. Cells were washed and harvested in PBS and

16   lysed in p300 lysis buffer. 3 mg of total cell lystate were precleared with 3% protein A-sepharose

17   beads and immunoprecipitated as described previously (41).

18

19   **Alkaline phosphatase assay.** Cell monolayers were rinsed in PBS, fixed in 100% methanol,

20   rinsed with PBS, then overlaid with 1.5 ml of 0.15 mg/ml BCIP (Sigma) plus 0.3 mg/ml NBT

21   (Promega, Madison, WI) for thirty minutes and rinsed again with PBS.

22

23

24   **RNA analysis and Probes.** Total cell RNA was prepared at designated times post-induction

25   using Trizol Reagent (GibcoBRL, Grand Island, NY) according to manufacturer's

26   recommendations. RNA in serial 10-fold dilutions (10 μg, 1 μg, 0.1 μg) was applied to

27   nitrocellulose BA85 in a slot-blotting apparatus (Schleicher and Schuell), and crosslinked by UV

1   irradiation. ^{32}P-labelled probes were prepared using a random primed labeling kit (Boehringer-

2   Mannheim). The histone H4 probe was described previously (35). Plasmid pGB.GAPDH was

3   constructed from MC3T3-E1 cell RNA by generating an RT-PCR fragment using primers

4   (5'ACTTTGTCAAGCTCATTTCC-3') and (5'-TGCAGCGAACTTTATTGATG-3') corresponding to

5   the murine glyceraldehyde-3-phosphate dehydrogenase cDNA sequence, and subcloning the

6   resulting PCR fragment into the TA cloning vector, pCR2.1 (Invitrogen, Carlsbad, CA).

7

8

9   **Antibodies.**   The p270-specific monoclonal antibody PSG3 and the ARID1B-specific

10   monoclonal antibody KMN1 have been described previously (34). A peptide used to generate a

11   BAF155-specific monoclonal antibody, DXD7, (34) also gave rise to a distinct BAF155-reactive

12   monoclonal antibody, designated DXD12, which cross-reacts with BAF170, and was used here.

13   An SV40 Tag-specific monoclonal antibody, mAb 419 (obtained from Ed Harlow), was used as a

14   negative control. Commercially purchased antibodies include the p21$^{CIP1/WAF1/SDI}$-specific antibody

15   (BD Biosciences, San Jose, CA) and the hsc70-specific antibody (Stressgen, San Diego, CA), as

16   well as antibodies of the following specificities obtained from Santa Cruz: cyclin A (C-19, sc-596),

17   cyclin B2 (N-20, sc-5235), and cyclin C (H-184, sc-5610). Rabbit polyclonal serum was raised

18   against the cdc2-G6 peptide sequence CDNQIKKM.

19

20   **Kinase assays.**   The cdc2-dependent kinase assays were performed as described previously

21   (42), using cdc2-specific immunoprecipitation complexes from 1 mg of total cell lysate and

22   histone H1 as exogenous substrate.

23

24   **DNA synthesis assay.**   Induced cells were labeled with ^{3}H-thymidine (Perkin-Elmer) (5 μCi/ml of

25   culture medium) in one hour pulses at the times post-induction indicated in the text , lysed in 0.3

26   M NaOH and assayed for trichloracetic acid (TCA)-precipitable counts as described previously

27   (43).

1

2   **Virus Infection.**   The generation and culture of the E1A-inactivated 9S adenovirus (used here

3   as a negative control) has been described previously (44). A stock of p21 expression virus

4   (Ad5CMVp21) (45) was provided by Judit Garriga, (Fels Institute, Temple University School of

5   Medicine, Philadelphia PA). MC3T3-E1 cells were infected at a multiplicity of infection of 25

6   plaque forming units per cell.

1   **RESULTS**

2

3   **Generation of p270-deficient and ARID1B-deficient MC3T3-E1-derived cell lines.**

4   Potential interfering oligonucleotide sequences were tested in a pSUPER-derived system

5   by standard protocols. Effective sequences were identified and introduced by stable integration

6   from the plasmid vector into the MC3T3-E1 line. Depletion was monitored by Western blotting

7   with p270-specific and ARID1B-specific monoclonal antibodies. Ten independent lines with

8   reduced expression of each target were selected, amplified and stored. As a control, vector-only

9   lines were selected and amplified in parallel. In each transfection, colonies appeared at similar

10   frequencies and showed essentially the same doubling time in normal growth medium as parental

11   cells. A representative p270 knockdown line (MC.p270.KD.AA2) and ARID1B knockdown line

12   (MC.1B.KD.CA6B) are each shown in **Fig. 1A**. The depleted lines (lane 3 and lane 6) show weak

13   p270 or ARID1B signals respectively in comparison with the parental line or a clonal line isolated

14   after transfection with the empty vector (lanes 1, 2, 4, and 5). The blots were additionally probed

15   with a monoclonal antibody that recognizes the closely related BAF155 and BAF170 mammalian

16   SWI/SNF complex subunits. Expression of these subunits is similar in each line. The overall

17   integrity of the complexes in the p270-depleted cells was verified by immunoprecipitation of ^{35}S-

18   labeled cell lysates with the BAF155-reactive antibody (**Fig. 1B**). Maintenance of expression of

19   the alternative ARID family protein in the conversely depleted cells was confirmed by

20   immunoprecipitation with p270-specific and ARID1B specific monoclonal antibodies (**Fig. 1C**). All

21   selected lines showed a similar degree of depletion. ^{35}S-methionine pulse labeling indicates that

22   new synthesis of each protein is reduced about ten-fold in the knockdown lines. As a further

23   probe for the overall integrity of the complexes in the knockdown lines, the remaining ARID family

24   product was removed by immune depletion of the cell lysates before immunoprecipitation with the

25   BAF155-reactive antibody (**Fig. 1D**). The results confirm that complex assembly is stable in the

26   absence of either subunit.

27

1   **p270-depleted and ARID1B-depleted cells both show impaired induction of the tissue-**

2   **specific marker alkaline phosphatase.**

3   MC3T3-E1 cells continue to proliferate for several days after induction with ascorbic acid,

4   then undergo cell cycle arrest at about day 3 post-induction. Expression of the earliest

5   differentiation marker, alkaline phosphatase, can be detected at this time. The knockdown lines

6   were tested for induction of alkaline phosphatase in an *in situ* enzyme assay scored by color

7   development (positive cells stain purple-black).   Three independent knockdown lines from each

8   series were tested, and all showed severe impairment of alkaline phosphatase induction; vector-

9   only lines behaved like the parental line.. In multiple independent experiments both series of

10   knockdown lines showed severely reduced induction of alkaline phosphatase at every point

11   tested, both early and late. Representative experiments at day 3 and day 14 are shown in **Fig. 2**.

12   These results indicate that the level of depletion achieved for each target is functionally

13   significant, and that both ARID-containing products are required for normal onset of

14   differentiation. We have used these knockdown lines to explore the role of each protein

15   specifically in cell cycle arrest functions.

16

17

18   **p270-deficient cells fail to undergo normal cell cycle arrest.**

19   The effect of p270 deficiency *versus* ARID1B deficiency was tested on specific cell cycle

20   arrest functions. We have previously used a gene array approach to identify many of the changes

21   in gene expression that occur in MC3T3-E1 cells as they procede through the differentiation

22   program (35). Expression was assayed on the arrays at days 0, 3, 7, and later times post-

23   induction. Between day 0 and day 7 a number of changes occurred that corresponded with the

24   shut-down of cell cycle activity. Among the most prominent was induction of the cell cycle

25   inhibitor, p21$^{Cip1/Waf1.}$ Several-fold induction of p21 was apparent by day 3 post-induction. When

26   this response was tested here at the protein level, a similar pattern of induction was apparent in

27   the parental line, but p21 expression was not induced in p270-depleted cells. A representative

1   Western blot depicting results from the MC.p270.KD.CA6 line is shown in **Fig. 3A**. The same

2   pattern was seen with the MC.p270.KD.AA2 and MC.p270.KD.DD2 lines (not shown). In contrast

3   to the p270-depleted cell lines, the ARID1B-depleted lines showed no impairment of p21

4   induction. Results from a representative line (MC.1B.KD.FD2) are shown in Fig. 3A; the same

5   result was observed with the MC.1B.KD.CA6B and MC.1B.KD.JD6 lines. As a loading control,

6   the blots were also probed with an antibody reactive against the constitutive form of the 70 kDa

7   heat shock protein, hsc70. The hsc70 signal was similar in all lanes.

8

9   The gene array also indicated down-regulation of several cyclins as the MC3T3-E1 cells

10   enter growth arrest. These responses were probed here with the MC.p270.KD.CA6 line and the

11   MC.1B.KD.FD2 line (**Fig. 3B**). The array showed down-regulation of B-type cyclins, particularly

12   cyclin B2, by day 3 post-induction. Consistent with the RNA signals, a decreased level of cyclin

13   B2 was apparent in the parental cells by day 4 on the Western blot. However, p270-depleted

14   cells again failed to show the parental response; levels of cyclin B2 remained high. Cyclin C was

15   also sharply down-regulated by day 3 in the gene array probe. Consistent with the RNA signal,

16   down-regulation of cyclin C in the parental cells was clear by day 2 in the Western blot. However,

17   cyclin C levels were unaffected by the induction protocol in the p270-depleted cells. The gene

18   array also indicated down-regulation of cyclin A, but this response was delayed relative to the

19   response patterns of cyclins B and C. On the array, a decreased cyclin A signal was apparent at

20   day 7 post-induction, but not at day 3. The protein probe shows a slight decrease in the cyclin A

21   level in the parental cells by day 6, and no detectable decrease in the p270-depleted cells. The

22   cdc2 kinase gene was not represented on the array, but expression of this gene is of interest as a

23   known E2F target subject to pRb-mediated repression (reviewed in 46). The cdc2 protein product

24   was assayed by Western blotting (**Fig. 3B**), which shows a sharp decrease in protein levels in the

25   parental cells by day 4 post-induction, and no detectable decrease in p270-depleted cells by day

26   6. In each of these assays, the ARID1B-depleted cells behaved indistinguishably from the

27   parental cells.

13

1

2   A well-characterized marker of proliferation state in differentiating osteoblasts is histone

3   H4 expression (reviewed in 39). Expression of this marker declines dramatically as the cells

4   arrest after induction. This response was also compared in parental MC3T3-E1 and the

5   knockdown cells. Consistent with other markers of cell cycle activity, the histone H4 signal

6   decreased sharply in the parental cells and the ARID1B knockdown line, but remained high in the

7   p270 knockdown line (**Fig. 4**).

8

9   The gene expression patterns that accompany induction to the differentiation phenotype

10   in parental MC3T3-E1 cells imply that a sharp decline in cyclin dependent kinase activity would

11   occur by day 4. The activity of cdc2-associated complexes was assayed here directly. The

12   kinase activity shows a sharp decline by day 4 post-induction in the parental cells and the

13   ARID1B-depleted cells, but no decline was detectable in the p270-depleted cells, even at day 6.

14   A representative kinase assay performed with the MC.p270.KD.DD2 line and the MC.1B.KD.FD2

15   line is shown in **Fig. 5A**. Results from independent experiments with three different p270-

16   depleted lines (MC.p270.KD.AA2, MC.p270.KD.CA6 and MC.p270.KD.DD2) and three different

17   ARID1B-depleted lines (MC.1B.KD.FD2, MC.1B.KD.JD6, and MC.1B.KD.CA6B) were quantified

18   on a phosphoimager and the averages are shown graphically in **Fig. 5B**.

19

20   The cell cycle status of the cells was probed directly by assessing the rate of ^{3}H-

21   thymidine incorporation over several days post-induction. Parental cells, p270-depleted cells,

22   and ARID1B-depleted cells were plated and induced in parallel. At 24 hour intervals, ^{3}H-

23   thymidine was added to the culture medium for one hour, after which the labeled cells were

24   harvested and assayed for incorporation of the isotope. The results shown for the parental cells

25   are the averages of three independent platings. The results shown for the p270-depleted cells

26   are the averages from three different knockdown lines (MC.p270.KD.AA2, MC.p270.KD.CA6 and

27   MC.p270.KD.DD2), as are the results shown for the ARID1B-depleted lines (MC.1B.KD.FD2,

14

1   MC.1B.KD.JD6, and MC.1B.KD.CA6B). The parental cells show a sharp decline in the rate of ^{3}H-

2   thymidine incorporation by day 4 post-induction, indicating a shut-down of DNA synthesis

3   consistent with cell cycle arrest. The same pattern is seen in the ARID1B-depleted lines. In

4   contrast, the rate of ^{3}H-thymidine incorporation decreases only slightly in the p270-depleted cells,

5   indicating continued DNA synthesis and failure to undergo normal cell cycle arrest (**Fig. 5C**).

6

7   These results identify the p270 subunit as critical for normal cell cycle arrest. In each of

8   the cell cycle assays described here, cloned cell lines containing a functional ARID1B-targeted

9   siRNA sequence behaved exactly like the parental line, indicating that the failure of p270-

10   depleted cells to undergo a normal cell cycle arrest response is a specific effect of p270-

11   deficiency.

12

13

14   **Induction of p21 and repression of E2F-responsive promoters are independent events that**

15   **each require p270.**

16   Previous studies have used differential expression of BRG1 and hSNF5 to probe the role

17   of SWI/SNF-related complexes in cell cycle arrest. Several studies, often using cloned rather

18   than endogenous promoters, found BRG1 enhances pRb-mediated repression of E2F-responsive

19   genes, and suggest that SWI/SNF subunits are associated with pRb in repressor complexes (23;

20   25; 47; 48). Decreased expression of endogenous E2F-responsive gene products such as cyclin

21   A and cdc2 was generally apparent at the protein level when BRG1 expression was restored to

22   naturally deficient tumor cell lines (20; 23), but in another study the effects were modest or cell

23   line specific, at least at the RNA level (19). An upstream effect with the potential to activate pRb

24   was seen consistently; exogenous expression of BRG1 in BRG1-deficient tumor cell lines results

25   in a sharp increase in p21 expression with most other cell cycle inhibitors, including p16[ink4a],

26   remaining relatively unaffected (19; 20). Expression of hSNF5 in malignant rhabdoid tumor

27   (MRT) lines has likewise been linked with decreased levels of E2F-responsive gene products

1   such as cyclin A (21; 22; 49).   Restoring hSNF5 to MRT cells does not alter p21 expression, but

2   up-regulates the p16^{ink4a} cell cycle inhibitor (17; 22; 46) in contrast to the effect of BRG1. The

3   difference may be a function of cell type rather than subunit effect, however, because siRNA-

4   mediated depletion of hSNF5 in HeLa (cervical carcinoma) or MG63 (osteosarcoma) cells causes

5   a sharp decrease in p21 levels, with no effect on p16 (20).

6

7   The failure of p270-depleted cells to induce p21 could theoretically account for all of the

8   cell cycle arrest defects observed in these lines. Repression of E2F-responsive genes and the

9   downstream effects of this repression might be impaired indirectly if cyclin-dependent kinase

10   activity is not appropriately inhibited, leaving targets such as pRb phosphorylated, inactive, and

11   unable to mediate repression of E2F-responsive genes. The effect of the complexes on the p21

12   promoter is apparently direct, as two studies have demonstrated the presence of BRG1 at the

13   p21 promoter, although a target element in the promoter could not be established (19; 20). A key

14   mechanistic question that remains unclear is whether down-regulation of E2F responsive genes

15   requires the action of the complexes independently of their effects on p21 levels. Chromatin

16   association assays have limited usefulness for these studies because the complexes associate

17   with chromatin widely.   This question was therefore addressed here genetically by introducing

18   exogenous expression of p21 in the p270-depleted cells simultaneously with the differentiation

19   signal.

20

21   Parental and p270-depleted cells were infected in parallel at the time of ascorbic acid

22   induction with an adenovirus vector expressing p21 or with a negative control virus containing an

23   inactivated E1A gene. The cells were monitored for DNA synthesis as described above.

24   Parental cells that received the p21 expression construct underwent accelerated shutdown of

25   DNA synthesis; ^{3}H-thymidine incorporation was severely repressed by day 2 post-induction (**Fig.**

26   **6 panel A**). Uninfected cells, or cells infected with the negative control virus, showed the same

27   kinetics seen in Fig. 6, i.e., without exogenous p21, DNA synthesis in the parental line remained

1   high at day 2; a severe decrease was not seen until day 4. In the p270-depleted cells, without

2   exogenous expression of p21, DNA synthesis remains high at least until day 6. However,

3   exogenous expression of p21 caused DNA synthesis to shut down with the same rapid kinetics

4   seen in the parental cells (**Fig. 6 panel B**). (Exogenous expression of p21 was verified by

5   Western blotting, shown in the left hand lanes of **Fig. 6 panel C**; normal induction of p21 in the

6   parental cells and the failure of p21 induction in the p270 depleted cells can be seen in the right

7   hand, 9S-infected, control lanes). The rapid down-regulation of DNA synthesis associated with

8   exogenous p21 expression was expected even in the p270-depleted cells because p21-induced

9   inhibition of cyclin dependent kinase activity is expected to result in the inactivation of essential

10 DNA replication factors. What is of special interest here is the status of E2F-responsive

11 products. This was monitored by Western blotting for the representative E2F-responsive gene

12 products cdc2, cyclin A, and cyclin B2. Expression of cyclin C was also examined. (**Fig. 6 panel**

13 **C**). The results show that expression levels of cdc2 and the cyclins remain high in the p270-

14 depleted cells despite exogenous expression of p21, indicating that regulation at the p21

15 promoter and at the E2F-responsive promoters each independently requires the function of the

16 chromatin remodeling complexes, and of p270 specifically, during differentiation-associated cell

17 cycle arrest. The expected repression of the cdc2-associated kinase activity in p21-expressing

18 p270-knockdown cells, despite the maintenance of a high level of cdc2 expression, was verified

19 in a kinase assay (**Fig. 6 panel D**).

20

21

22

23

24

1 **DISCUSSION**

2

3   The work described here identifies the p270 subunit of mammalian SWI/SNF-related

4 complexes as critical for normal cell cycle arrest in differentiating cells exiting the cell cycle. The

5 evidence of p270 deficiency in certain tumors and tumor cell lines (32), implied indirectly that

6 p270 plays a required role in the tumor suppressor activity of the complex(es). The results

7 presented here establish a specific biological basis for the clinical findings, demonstrating directly

8 that p270 is essential for both the induction of p21 and the repression of E2F responsive genes

9 such as cdc2 during differentiation-associated cell cycle arrest. Involvement in both activation

10 and repression is a feature of SWI/SNF-related complexes generally, presumably determined by

11 the spectrum of transactivators and repressors that recruit the complexes. The demonstration

12 that the complexes are required separately for regulation of at least two distinct steps in

13 proliferation control underscores the carcinogenic potential of cells that have lost function of a

14 required subunit.

15

16   These results are particularly significant because previous studies concerning the roles of

17 SWI/SNF complex components in expression of cell cycle markers have largely relied on re-

18 introduction of BRG1 or hSNF5 into tumor cell lines where they were lacking (e.g. 17; 19; 20; 22;

19 23; 25; 49), rather than monitoring the role of complex components in cells undergoing

20 physiological progression from a proliferative state to cell cycle arrest. The identification of p270

21 as a subunit required for cell cycle arrest *in vivo* is additionally significant, because unlike BRG1

22 and hSNF5, p270 is not among those subunits considered to form the "functional core" of the

23 complex(es). The concept of a functional core was based on the observation that BRG1 has a

24 relatively low level of enzyme activity when purified away from other subunits, and that a level of

25 remodeling and ATPase activity similar to that of the intact complex(es) can be reconstituted *in*

26 *vitro* by assembly of a subset of components consisting of BRG1, BAF170, BAF155, and hSNF5

27 (50). The *in vivo* requirement for p270 shows that it plays an essential role in the physiological

1   functions of the complex(es), regardless of whether it contributes directly to the overall level of

2   enzymatic activity.

3

4   A cell cycle arrest function has previously been ascribed to the BRG1 and hSNF5

5   components; however, those subunits do not distinguish between the BAF and PBAF complexes.

6   The p270–depletion phenotype constitutes the first formal evidence of a requirement for the BAF

7   complex series, as opposed to the PBAF complex, in cell cycle regulation. The data presented

8   here shed further light on the question of specificity among the several distinct configurations of

9   the BAF complexes. The alternative ATPase subunits, BRM and BRG1, may be partially

10   redundant in their ability to support cell cycle arrest when exogenously expressed (51), but the

11   most physiological experiments (3; 4) suggest strongly that BRG1-specific complexes and not

12   BRM-specific complexes are essential for this function. The present study indicates that p270-

13   containing complexes, but not ARID1B-containing complexes, are required for cell cycle arrest.

14   Logically, it appears that, of the four combinations made possible by these alternative subunits, it

15   is the BRG1 and p270 combination that plays the major role in cell cycle arrest. hSNF5 is not a

16   determinant of specificity between complexes, but its presence along with BRG1 and p270 is

17   required for the activities that the complex contributes to cell cycle arrest. These findings will help

18   to clarify targets for drug intervention therapies.

19

20   The essential biochemical activities contributed by the noncatalytic subunits have yet to

21   be determined. The amino acid sequence of hSNF5 gives little clue to the function of the protein,

22   which remains unknown, except that it can facilitate DNA end-joining *in vitro* (52). In p270, the

23   most obvious structural motif is the approximately 100 amino acid long ARID DNA binding

24   domain, but this large protein (2,285 amino acids) also contains potential protein-protein

25   interaction surfaces (27; 28; 33; 41) that may be more important for its specific function. ARID1B

26   contains an ARID domain that is 80% identical with p270 (alignments can be seen in 30 and 53),

27   and both proteins belong to a subclass of the ARID family that binds DNA without regard to

19

1   sequence specificity (34; 53). Thus, a likely scenario is that DNA binding is a function common to

2   both, while sequences outside the ARID determine specificity of function. Structure-function

3   analysis of both these ARID-containing subunits is in progress.

4

5   The effects linked individually with BRG1, hSNF5 and p270 are generally similar but not

6   identical. What is not clear is whether differences so far reported are due more to methodology

7   or cell type than to true differences in function between the subunits. Approaches based on re-

8   introduction of specific components into deficient cell tumor cell lines preclude the ability to

9   compare function within a single cell line or in non-transformed lines. The ability to knock down

10   expression is freeing investigators to study the role of the complexes in nontransformed cell lines.

11   BRG1 and BRM can be inhibited by dominant/negative forms with an inactivating mutation in the

12   ATP binding site. Dominant/negative inhibition of BRG1/BRM in NIH3T3 mouse fibroblasts

13   inhibited MyoD-dependent differentiation, but surprisingly did not inhibit concomitant cell cycle

14   arrest; p21 is induced in these conditions and its induction was likewise unaffected by expression

15   of the dominant/negative construct (54). Cell cycle arrest independent of SWI/SNF complex

16   activity may be a phenomenon specific to the function of MyoD however, because a similar

17   dominant/negative approach in BALB/c mouse fibroblasts was sufficient to inhibit C/EBPα-

18   induced cell cycle arrest severely (55). In the latter study siRNA-mediated depletion of hSNF5 or

19   BRM had the same effect on cell growth curves as expression of the dominant/negative

20   construct, but expression of individual genes was not assessed.

21

22   The system used here maintains the benefits of probing function in non-transformed

23   cells, and has the additional advantage of relying on an external induction signal to initiate

24   differentiation and cell cycle arrest functions rather than engineered over-expression of a single

25   gene such as MyoD or C/EBPα. The MC3T3-E1 cell system is amenable to knockdown studies

26   targeting each of a succession of subunits, and probing an array of extracellular signals. Further

27   studies are in progress.

1   ACKNOWLEDGEMENTS

2

1 **REFERENCES**

2 1. Martens JA, Winston F. Recent advances in understanding chromatin remodeling by Swi/Snf

3   complexes. Curr Opin Genet Dev 2003; 13:136-42.

4

5 2. Roberts CW, Orkin SH. The SWI/SNF complex--chromatin and cancer. Nat Rev Cancer 2004;

6   4:133-42.

7

8 3. Bultman S, Gebuhr T, Yee D, La Mantia C, Nicholson J, Gilliam A, Randazzo F, Metzger D,

9   Chambon P, Crabtree G, Magnuson T. A Brg1 null mutation in the mouse reveals functional

10   differences among mammalian SWI/SNF complexes. Mol Cell 2000; 6:1287-95.

11

12 4. Reyes JC, Barra J, Muchardt C, Camus A, Babinet C, Yaniv M. Altered control of cellular

13   proliferation in the absence of mammalian brahma (SNF2alpha). EMBO J 1998; 17:6979-91.

14

15 5. Fukuoka J, Fujii T, Shih JH, Dracheva T, Meerzaman D, Player A, Hong K, Settnek S, Gupta

16   A, Buetow K, Hewitt S, Travis WD, Jen J. Chromatin remodeling factors and BRM/BRG1

17   expression as prognostic indicators in non-small cell lung cancer. Clin Cancer Res 2004;

18   10:4314-24.

19

20 6. Reisman DN, Sciarrotta J, Wang W, Funkhouser WK, Weissman BE. Loss of BRG1/BRM in

21   human lung cancer cell lines and primary lung cancers: correlation with poor prognosis.

22   Cancer Res 2003; 63:560-6.

23

24 7. Wong AK, Shanahan F, Chen Y, Lian L, Ha P, Hendricks K, Ghaffari S, Iliev D, Penn B,

25   Woodland AM, Smith R, Salada G, Carillo A, Laity K, Gupte J, Swedlund B, Tavtigian SV,

26   Teng DH, Lees E. BRG1, a component of the SWI-SNF complex, is mutated in multiple

27   human tumor cell lines. Cancer Res 2000; 60:6171-7.

8. Guidi CJ, Sands AT, Zambrowicz BP, Turner TK, Demers DA, Webster W, Smith TW, Imbalzano AN, Jones SN. Disruption of Ini1 leads to peri-implantation lethality and tumorigenesis in mice. Mol Cell Biol 2001; 21:3598-603.

9. Klochendler-Yeivin A, Fiette L, Barra J, Muchardt C, Babinet C, Yaniv M. The murine SNF5/INI1 chromatin remodeling factor is essential for embryonic development and tumor suppression. EMBO Rep 2000; 1:500-6.

10. Roberts CW, Galusha SA, McMenamin ME, Fletcher CD, Orkin SH. Haploinsufficiency of Snf5 (integrase interactor 1) predisposes to malignant rhabdoid tumors in mice. Proc Natl Acad Sci USA 2000; 97:13796-800

11. Biegel JA, Tan L, Zhang F, Wainwright L, Russo P, Rorke LB. Alterations of the hSNF5/INI1 gene in central nervous system atypical teratoid/rhabdoid tumors and renal and extrarenal rhabdoid tumors. Clin Cancer Res 2002; 8:3461-7

12. Uno K, Takita J, Yokomori K, Tanaka Y, Ohta S, Shimada H, Gilles FH, Sugita K, Abe S, Sako M, Hashizume K, Hayashi Y. Aberrations of the hSNF5/INI1 gene are restricted to malignant rhabdoid tumors or atypical teratoid/rhabdoid tumors in pediatric solid tumors. Genes Chromosomes Cancer 2002; 34:33-41.

13. Versteege I, Sevenet N, Lange J, Rousseau-Merck MF, Ambros P, Handgretinger R, Aurias A, Delattre O. Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. Nature 1998; 394:203-206.

1   14. Biegel JA, Zhou JY, Rorke LB, Stenstrom C, Wainwright LM, Fogelgren B. Germ-line and

2   acquired mutations of INI1 in atypical teratoid and rhabdoid tumors. Cancer Res 1999; 59:74-

3   9.

4

5   15. Sevenet N, Sheridan E, Amram D, Schneider P, Handgretinger R, Delattre O. Constitutional

6   mutations of the hSNF5/INI1 gene predispose to a variety of cancers. Am J Hum Genet 1999;

7   65:1342-8.

8

9   16. Taylor MD, Gokgoz N, Andrulis IL, Mainprize TG, Drake JM, Rutka JT. Familial posterior

10   fossa brain tumors of infancy secondary to germline mutation of the hSNF5 gene. Am J Hum

11   Genet 2000; 66:1403-6.

12

13   17. Betz BL, Strobeck MW, Reisman DN, Knudsen ES, Weissman BE. Re-expression of

14   hSNF5/INI1/BAF47 in pediatric tumor cells leads to G1 arrest associated with induction of

15   p16ink4a and activation of RB. Oncogene 2002; 21:5193-203.

16

17   18. Dunaief JL, Strober BE, Guha S, Khavari PA, Alin K, Luban J, Begemann M, Crabtree GR,

18   Goff SP. The retinoblastoma protein and BRG-1 form a complex and cooperate to induce cell

19   cycle arrest. Cell 1994; 79:119-130.

20

21   19. Hendricks KB, Shanahan F, Lees E. Role for BRG1 in cell cycle control and tumor

22   suppression. Mol Cell Biol 2004; 24:362-76.

23

24   20. Kang H, Cui K, Zhao K. BRG1 controls the activity of the retinoblastoma protein via

25   regulation of p21CIP1/WAF1/SDI. Mol Cell Biol 2004; 24:1188-99.

26

1   21. Medjkane S, Novikov E, Versteege I, Delattre O. The tumor suppressor hSNF5/INI1

2   modulates cell growth and actin cytoskeleton organization. Cancer Res 2004; 64:3406-13.

3

4   22. Oruetxebarria I, Venturini F, Kekarainen T, Houweling A, Zuijderduijn LM, Mohd-Sarip A,

5   Vries RG, Hoeben RC, Verrijzer CP. p16INK4a is required for hSNF5 chromatin remodeler-

6   induced cellular senescence in malignant rhabdoid tumor cells. J Biol Chem 2004; 279:3807-

7   16.

8

9   23. Strobeck MW, Knudsen KE, Fribourg AF, DeCristofaro MF, Weissman BE, Imbalzano AN,

10   Knudsen ES. BRG-1 is required for RB-mediated cell cycle arrest. Proc Natl Acad Sci USA

11   2000; 97:7748-53.

12

13   24. Versteege I, Medjkane S, Rouillard D, Delattre O. A key role of the hSNF5/INI1 tumour

14   suppressor in the control of the G1-S transition of the cell cycle. Oncogene 2002; 21:6403-12.

15

16   25. Zhang HS, Gavin M, Dahiya A, Postigo AA, Ma D, Luo RX, Harbour JW, and Dean DC. Exit

17   from G1 and S phase of the cell cycle is regulated by repressor complexes containing HDAC-

18   Rb-hSWI/SNF and Rb-hSWI/SNF. Cell 2000; 101:79-89.

19

20   26. Choi JY, Lee BH, Song KB, Park RW, Kim IS, Sohn KY, Jo JS, Ryoo HM. Expression

21   patterns of bone-related proteins during osteoblastic differentiation in MC3T3-E1 cells. J Cell

22   Biochem 1996; 61:609-18.

23

24   27. Dallas PB, Pacchione S, Wilsker D, Bowrin V, Kobayashi R, Moran E. The human SWI/SNF

25   complex protein, p270, is an ARID family member with nonsequence -specific DNA binding

26   activity. Mol Cell Biol 2000; 20:3137-3146.

27

1   28. Nie Z, Xue Y, Yang D, Zhou S, Deroo BJ, Archer TK, Wang W. A specificity and targeting

2   subunit of a human SWI/SNF family-related chromatin-remodeling complex. Mol Cell Biol

3   2000; 20:8879-8888.

4

5   29. Wilsker D, Patsialou A, Dallas PB, Moran E. ARID proteins: A new family of DNA binding

6   proteins implicated in the control of cell growth, differentiation, and development. Cell Growth

7   & Differentiation 2002; 13:95-106.

8

9   30. Wilsker D, Probst L, Wain HM, Maltais L, Tucker PW, Moran E. Nomenclature of the ARID

10   family of DNA binding proteins. Genomics 2005; *in press.*

11

12   31. DeCristofaro MF, Betz BL, Rorie CJ, Reisman DN, Wang W, Weissman BE. Characterization

13   of SWI/SNF protein expression in human breast cancer cell lines and other malignancies. J

14   Cell Physiol 2001; 186:136-145.

15

16   32. Wang X, Nagl NG Jr, Flowers S, Zweitzig DR, Dallas PB, Moran E. Expression of p270

17   (ARID1A), a component of human SWI/SNF complexes, in human tumors. Int J Cancer 2004;

18   112:636-642.

19

20   33. Inoue H, Furukawa T, Giannakopoulos S, Zhou S, King DS, Tanese N. Largest subunits of

21   the human SWI/SNF chromatin remodeling complex promote transcriptional activation by

22   steroid hormone receptors. J Biol Chem 2002; 277:41674-85.

23

24   34. Wang X, Nagl NG Jr, Van Scoy M, Pacchione S, Yaciuk P, Dallas PB, Moran E. Two related

25   ARID family proteins are alternative subunits of human SWI/SNF complexes. Biochemical

26   Journal 2004; 383:319-325.

27

1   35. Beck GR Jr, Zerler B, Moran E. Gene array analysis of osteoblast differentiation. Cell

2   Growth and Differentiation 2001; 12:61-83.

3

4   36. Quarles LD, Yohay DA, Lever LW, Caton R, Wenstrup RJ. Distinct proliferative and

5   differentiated stages of murine MC3T3-E1 cells in culture: an in vitro model of osteoblast

6   development. J Bone Miner Res 1992; 7:683-92.

7

8   37. Raouf A, Seth A. Ets transcription factors and targets in osteogenesis. Oncogene 2000;

9   19:6455-63.

10

11   38. Stein GS, Lian JB. Molecular mechanisms mediating proliferation/differentiation

12   interrelationships during progressive development of the osteoblast phenotype. Endocr Rev

13   1993; Aug;14:424-42.

14

15   39. Stein GS, Lian JB, Stein JL, Van Wijnen AJ, Montecino M. Transcriptional control of

16   osteoblast growth and differentiation. Physiol Rev 1996; 76:593-629.

17

18   40. Brummelkamp TR, Bernards R, Agami R. A system for stable expression of short interfering

19   RNAs in mammalian cells. Science 2002; 296:550-3.

20

21   41. Dallas PB, Cheney IW, Liao D, Bowrin V, Byam W, Pacchione S, Kobayashi R, Yaciuk P,

22   Moran E. p300/CREB binding protein-related protein p270 is a component of mammalian

23   SWI/SNF complexes. Mol Cell Biol 1998; 18:3596-603.

24

25

1   42. Wang HG, Draetta G, Moran E. E1A induces phosphorylation of the retinoblastoma protein

2   independently of direct physical association between the E1A and retinoblastoma products.

3   Mol Cell Biol 1991; 11:4253-65.

4

5   43. Zerler B, Roberts RJ, Mathews MB, Moran E. Different functional domains of the adenovirus

6   E1A gene are involved in regulation of host cell cycle products. Mol Cell Biol 1987; 7:821-9.

7

8   44. Zerler B, Moran E, Maruyama K, Moomaw J, Grodzicker T, and Ruley HE. Adenovirus E1A

9   coding sequences that enable ras and pmt oncogenes to transform cultured primary cells. Mol

10   Cell Biol 1986; 6:887-99.

11

12   45. Prabhu NS, Blagosklonny MV, Zeng YX, Wu GS, Waldman T, El-Deiry WS. Suppression of

13   cancer cell growth by adenovirus expressing p21(WAF1/CIP1) deficient in PCNA interaction.

14   Clin Cancer Res 1996; 2:1221-9.

15

16   46. Muller H, Helin K. The E2F transcription factors: key regulators of cell proliferation. Biochem

17   Biophys Acta 2000; 1470:M1-12.

18

19   47. Trouche D, Le Chalony C, Muchardt C, Yaniv M, Kouzarides T. RB and hbrm cooperate to

20   repress the activation functions of E2F1. Proc Natl Acad Sci USA 1997; 94:11268-73.

21

22   48. Wang S, Zhang B, Faller DV. Prohibitin requires Brg-1 and Brm for the repression of E2F and

23   cell growth. EMBO J 2002; 21:3019-28.

24

25   49. Reincke BS, Rosson GB, Oswald BW, Wright CF. INI1 expression induces cell cycle arrest

26   and markers of senescence in malignant rhabdoid tumor cells. J Cell Physiol 2003; 194:303-

27   13.

1

2   50. Phelan ML, Sif S, Narlikar GJ, Kingston RE. Reconstitution of a core chromatin remodeling

3   complex from SWI/SNF subunits. Mol Cell 1999; 3:247-53.

4

5   51. Reisman DN, Strobeck MW, Betz BL, Sciariotta J, Funkhouser W Jr, Murchardt C, Yaniv M,

6   Sherman LS, Knudsen ES, Weissman BE. Concomitant down-regulation of BRM and BRG1

7   in human tumor cell lines: differential effects on RB-mediated growth arrest vs CD44

8   expression. Oncogene 2002; 21:1196-1207.

9

10   52. Kalpana GV, Marmon S, Wang W, Crabtree GR, Goff SP. Binding and stimulation of HIV-1

11   integrase by a human homolog of yeast transcription factor SNF5. Science 1994; 266:2002-6.

12

13   53. Patsialou A, Wilsker D, Moran E. DNA-binding properties of ARID family proteins. Nucleic

14   Acids Res 2005; 33:66-80.

15

16   54. de la Serna IL, Roy K, Carlson KA, Imbalzano AN. MyoD can induce cell cycle arrest but not

17   muscle differentiation in the presence of dominant negative SWI/SNF chromatin remodeling

18   enzymes. J Biol Chem 2001; 276:41486-91.

19

20   55. Muller C, Calkhoven CF, Sha X, Leutz A. The CCAAT enhancer-binding protein alpha

21   (C/EBPalpha) requires a SWI/SNF complex for proliferation arrest. J Biol Chem 2004;

22   279:7353-8.

23

24

1 **FIGURE LEGENDS**

2

3

4 **Figure 1. Expression of p270 and ARID1B in MC3T3-E1-derived knockdown lines.** A. 100

5 μg of total cell lysate per lane was separated on 8% SDS-PAGE gels, transferred to PVDF

6 membrane, and probed with either p270-specific or ARID1B-specific monoclonal antibodies, and

7 with a BAF155/BAF170-reactive monoclonal antibody. B. Aliquots of ^{35}S-labeled cell lysates from

8 parental MC3T3-E1 cells (lanes 1 and 2) or MC.p270.KD.AA2 cells (lane 3) or MC.1B.KDCA6B

9 cells (lane 4) were immunoprecipitated with control antibody (lane 1) or a BAF155-reactive

10 monoclonal antibody (lanes 2 through 4). The identity of the higher molecular weight proteins

11 labeled at the right in the immune complex was verified directly by Western blotting. C. Aliquots

12 of ^{35}S-labeled cell lysates isolated as described in panel B were immunoprecipitated with p270-

13 specific or ARID1B-specific monoclonal antibodies, as indicated in the figure. D. (I): Aliquots of

14 ^{35}S-labeled ARID1B-knockdown cells were depleted of p270 by five successive

15 immunoprecipitations with a p270-specific mAb; lanes 1 through 5 show the autoradiogram signal

16 of p270 brought down by each successive immunoprecipitation. (II): Aliquots of ^{35}S-labeled p270

17 knockdown cells were depleted of ARID1B by five successive immunoprecipitations with an

18 ARID1B-specific mAb; lanes 6 through 10 show the radiogram signal of ARID1B brought down by

19 each successive immunoprecipitation. The fully depleted lysates from series I and II (sampled in

20 lanes 5 and 10, respectively) were then each immunoprecipitated with a BAF155-reactive

21 antibody (DXD12). Visualization of the immune complexes by autoradiography (lanes I and II)

22 shows the overall integrity of the complexes in the absence of both ARID family subunits.

23

24 **Figure 2. Differentiation phenotype of the knockdown cell lines.** Cells were assayed at day

25 0, day 3, and day 14 post-induction for alkaline phosphatase activity. Conversion of the substrate

26 to a purple color indicates activity of the enzyme. AA2, CA6, and DD2 are identification numbers

1   for three independently isolated p270-depleted lines.   JD6, FD2, and CA6B are identification

2   numbers for three independently isolated ARID1B-depleted lines.

3

4   **Figure 3. Expression of p21 and other cell cycle markers in p270-depleted and ARID1B-**

5   **depleted cells.** Parental, p270-depleted, and ARID1B-depleted MC3T3-E1 cells were harvested

6   at days 0, 2, 4 and 6 post-induction.   100 μg of total cell lysate per lane was separated on SDS-

7   PAGE gels, transferred to PVDF membrane, and probed sequentially with antibodies of each of

8   the specificities shown.

9

10   **Figure 4. Expression of histone H4 in p270-depleted cells.** A. Parental, p270-depleted, and

11   ARID1B-depleted MC3T3-E1 cells were harvested at days 0 and 3 post-induction. Total cell RNA

12   was applied to nitrocellulose and probed with a histone H4-specific probe and a GAPDH-specific

13   probe included as a loading control. B.   Quadruplicate samples were probed for histone H4

14   expression as described in panel A, quantified by phosphoimaging, normalized to GAPDH, and

15   plotted relative to day 0. The error bars indicate average deviation.

16

17   **Figure 5. cdc2/CDK1-associated kinase activity and DNA synthesis activity in p270-**

18   **depleted cells.** Parental, p270-depleted, and ARID1B-depleted MC3T3-E1 cells were harvested

19   at days 0, 2, 4 and 6 post-induction.   cdc2-specific immune complexes were isolated from 1 mg

20   of total cell lystate, and incubated with γ-^{32}P-ATP and histone H1 as exogenous substrate.   A.

21   The reactions were separated on 15% SDS-PAGE gels and visualized by fluorography. B.

22   Reactions performed as described in panel A were quantified by phosphoimaging. Results from

23   three independent experiments with independently isolated cell lines of each knockdown series

24   were averaged, and plotted as kinase activity relative to day 0. The average deviation at each

25   point is indicated by error bars.   The solid line indicates kinase activity in the parental cells; the

26   dashed line indicates activity in the p270-depleted cells; the dotted line indicates activity in the

27   ARID1B-depleted cells. C. Parental, p270-depleted, and ARID1B-depleted MC3T3-E1 cells

1   were labeled with ³H-thymidine in one hour pulses at days 0, 2, 4 and 6 post-induction, and

2   assayed for trichloracetic acid (TCA)-precipitable counts.   Results from three independent

3   platings of the parental cell line, and from three independently isolated cell lines from each

4   knockdown series were averaged within their respective groups and plotted as CPM

5   incorporated. The average deviation at each point is indicated by error bars. The striped bar

6   indicates incorporation in the parental cells; the solid bar indicates incorporation in the p270-

7   depleted cells. The open bar indicates incorporation in the ARID1B-depleted cells.

8

9   **Figure 6. Effect of exogenous expression of p21.** Parental and p270-depleted cells were

10   infected in parallel at the time of ascorbic acid induction (day 0) with an adenovirus vector

11   expressing p21 or with a negative control virus containing an inactivated E1A gene (9S). A and

12   B: Cells were assessed for DNA synthesis activity monitored by ³H-thymidine incorporation; the

13   curves are the averages and average deviation of results obtained with three independent

14   platings of the parental line in parallel with three independent p270 knockdown lines. C: Levels of

15   cdc2, p21, cyclins B2, A, and C, as well as hsc70 were probed by Western blotting in parental

16   and p270 knockdown cells infected with the p21-expressing virus or the control (9S) virus. D:

17   cdc2-associated kinase activity was assayed in parental and p270 knockdown cells infected with

18   the p21-expressing virus or the control (9S) virus.   The graphs show the averages and average

19   deviations from triplicate platings of the parental line and three different p270 knockdown lines;

20   the gel shows a representative result.
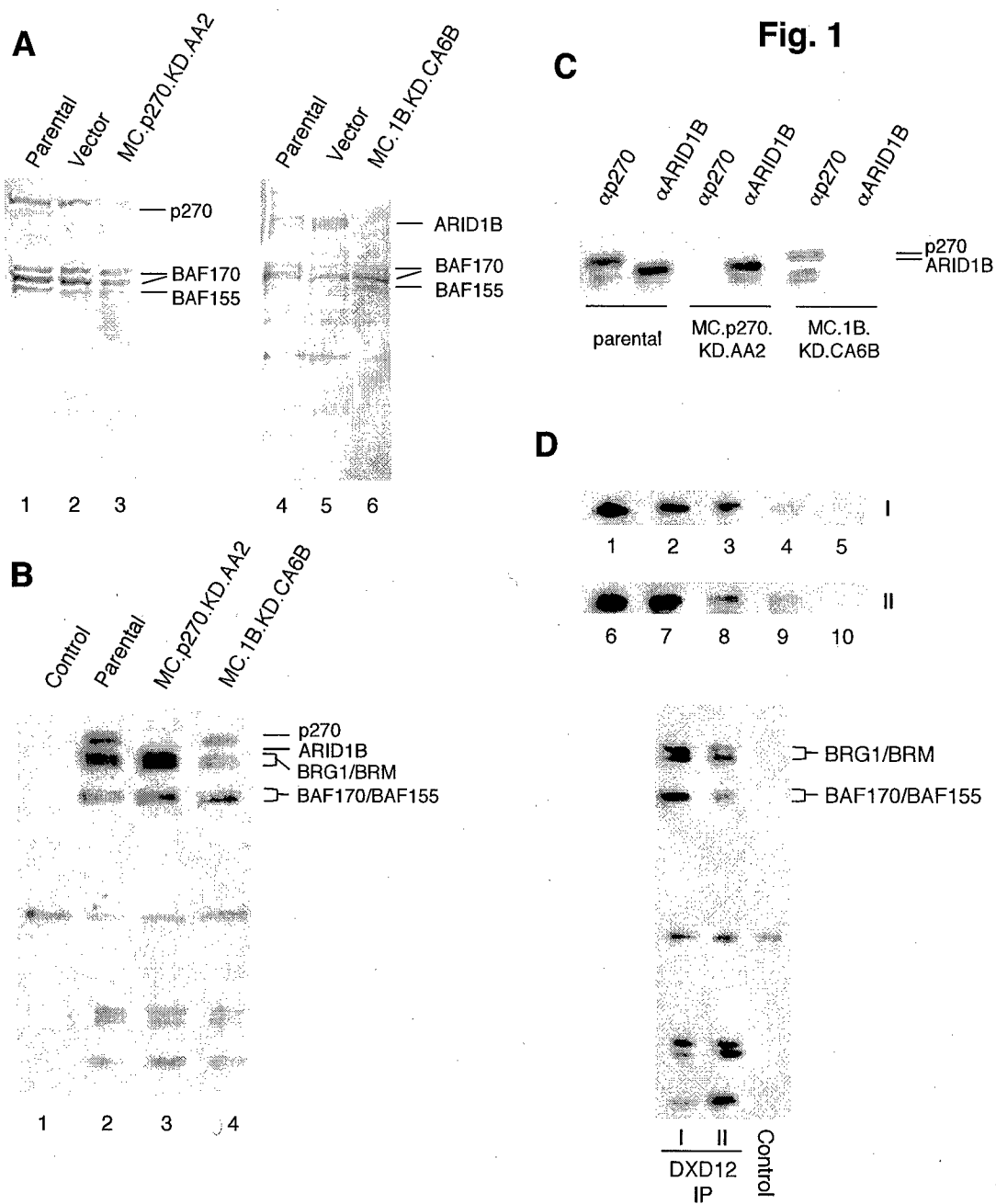
21

22

23

24

25

Fig. 1

A

Parental | Vector | MC.p270.KD.AA2 | Parental | Vector | MC.1B.KD.CA6B

— p270
— ARID1B
BAF170
BAF155

1 2 3   4 5 6

B

Control | Parental | MC.p270.KD.AA2 | MC.1B.KD.CA6B

— p270
ARID1B
BRG1/BRM
BAF170/BAF155

1 2 3 4

C

αp270 | αARID1B | αp270 | αARID1B | αp270 | αARID1B

p270
ARID1B

parental   MC.p270.
KD.AA2   MC.1B.
KD.CA6B

D

I
1 2 3 4 5

II
6 7 8 9 10

BRG1/BRM
BAF170/BAF155

I   II   Control
DXD12
IP

Fig. 2

Parental	Vector	AA2 CA6 DD2	JD6 FD2 CA6B
		MC.p270.KD.	MC.1B.KD.

34

A

Fig. 3

Parental | MC.p270 .KD.CA6 | MC.1B .KD.FD2

— p21

— hsc70

0 2 4 6 | 0 2 4 6 | 0 2 4 6 : Day

B

Parental | MC.p270 .KD.CA6 | MC.1B .KD.FD2

—Cyclin B2

—Cyclin C

—Cyclin A

—cdc2

0 2 4 6 | 0 2 4 6 | 0 2 4 6 : Day

A

Parental   MC.p270   MC.1B
.KD.AA2   .KD.CA6B

H4

GAPDH

Day: 0   3   0   3   0   3

B

Fig. 4

36

A

Parental | MC.p270 .KD.DD2 | MC.1B .KD.FD2

Day: 0 2 4 6 | 0 2 4 6 | 0 2 4 6

B

Kinase Activity Relative to Day 0(%)

p270.KD

Parental

ARID1B.KD

Day Post-Induction

C

⬚ Parental
▨ MC.p270.KD
☐ MC.1B.KD

CPM

Day: 0   2   4   6

Fig. 5

37

Fig. 6